

OM of: US-09-558-284-2 to: Geneml: * out_format : pfs

Date: Jul 22, 2001 9:26 PM

About: Results were produced by the GenCore software, version 4.5.
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+ip2n.model -DEV=xlp
-O=/cgn2_1/USPTO.spool/US09558284/funat_20072001.160145_19146/app_query.fasta_1.306
-DB=geneml -OPMT=fastap -SUFFIX=p2n.rge -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blotsum62
-TRAAS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pcu
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORExt -MINLEN=0 -MAXLEN=2000000000
-USER=US09558284.ecgn1_1_5075 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:

Query: US-09-558-284-2

Database: Geneml: *

Database sequences: 1344157

Database length: 856060004

Search time (sec): 1093.790000

score_list:

Sequence	Strid Orig	ZScore	EScore	len	Documentation
gb_pat1:A84558	+ 1269.00	2059.73	1.7e-106	758	A84558 Sequence 1 from Patent W
gb_pat1:A85030	+ 1269.00	2059.73	1.7e-106	758	A85030 Sequence 1 from Patent B
gb_pat1:SCU03107	+ 1269.00	2052.06	4.6e-106	1650	I 003107 Saccharomyces cerevisia
gb_pat1:ISC8179	+ 1269.00	2019.65	9.9e-104	44113	I 00062 Saccharomyces cerevisia
gb_sts1:CNS0616P	+ 544.00	879.41	5.9e-41	826	I AL359911 T3 end of clone AS0A0
gb_sts1:CNS072Y0	+ 459.00	739.48	5.9e-33	976	I AL426814 clone BA0A015C05 of 1
gb_sts1:CNS07BKB	+ 421.50	677.87	3.9e-29	1037	I AL437985 T7 end of clone BC0A
gb_pat1:CAC20C1	+ 417.00	635.04	3.9e-27	3768	I AL033391 C. albicans cosmid C
gb_pat1:YSCP450R	+ 357.00	562.10	4.5e-23	3112	I D13788 Yeast NADPH-cytochrome
gb_sts1:CNS07CNC	+ 333.50	534.90	1.5e-21	1017	I AL439410 T7 end of clone BD0A
gb_pat1:YSPR72W	+ 328.50	527.70	3.7e-21	925	I D50471 Yeast Rht2 mRNA, complet
gb_pat1:SCCE1	+ 328.50	516.30	1.6e-20	2940	I X73488 Saccharomyces cerevisia
gb_pat1:SCS610	+ 327.50	491.12	4.0e-19	37730	I Z38060 S. cerevisiae chromosome
gb_pat1:AX085442	+ 327.50	528.11	3.5e-21	753	I AX085442 Sequence 55 from Paten
gb_pat1:YSCOR2A	+ 327.50	525.62	4.8e-21	969	I D50469 Yeast Hor2 mRNA, complet
gb_pat1:AF270838	+ 327.50	511.80	5.8e-19	3931	I AF270838 Saccharomyces pastori
gb_pat1:SCCE592	+ 324.50	488.30	2.8e-19	42576	I U18813 Saccharomyces cerevisia
gb_sts1:CNS06P1	+ 324.50	525.94	4.6e-21	572	I AL409499 T3 end of clone AV0A0
gb_pat1:AE000318	+ 270.50	409.05	1.5e-14	10863	I AE000318 Escherichia coli K12
gb_pat1:AE005461	+ 270.50	409.04	1.5e-14	10866	I AE005461 Escherichia coli O15
gb_pat1:DB0861	+ 270.50	405.97	2.2e-14	14839	I D90861 E. coli genomic DNA, K
gb_pat1:D90860	+ 270.50	404.32	2.8e-14	17533	I D90860 E. coli genomic DNA, K
gb_pat1:AP002561	+ 270.50	376.42	9.9e-13	296827	I AP002561 Escherichia coli O3
gb_sts1:CNS06IMD	+ 269.00	430.21	2.0e-15	992	I AL400475 T3 end of clone AS0A0
gb_sts1:CNS073U2	+ 266.00	424.65	2.0e-15	1063	I AL428001 clone BA0A023B05 of
gb_pat1:RHEXGNS	+ 266.00	391.18	1.5e-13	31638	I Z79692 R. melioli exp gene cl
gb_pat1:SCU11	+ 255.00	377.18	9.0e-13	21327	I AL157953 Streptomyces coelic
gb_pat1:AF042322	+ 244.00	381.03	5.5e-13	2351	I AF042322 Emeritella nidulans m
gb_sts1:CNS06GJM	+ 222.00	358.03	1.0e-11	894	I AL397748 T3 end of clone AS0A0
gb_pat1:SCC8A	+ 222.00	319.40	1.5e-09	32274	I AL356892 Streptomyces coelic
gb_pat1:AP004003	+ 213.50	282.09	1.8e-07	349116	I AP003003 Mesorhizobium loti
gb_pat1:AE004633	+ 182.00	265.78	1.4e-06	10107	I AE004633 Pseudomonas aerugin
gb_sts1:CNS07A13	+ 158.00	249.75	1.1e-05	980	I AL437005 T7 end of clone BC0A0
gb_pat1:AC020875	+ 155.00	204.00	0.0040	6196	I AC020875 Mus musculus chromos
gb_sts1:AC020875	+ 153.00	242.21	3.0e-05	923	I AL423751 T7 end of clone AZ0A0
gb_pat1:SC32731	+ 151.00	215.57	0.0009	9881	I U32731 Haemophilus influenzae
gb_pat1:AE006146	+ 150.00	213.13	0.0012	10734	I AE006146 Pasteurella multocic
gb_sts1:CNS07ZAK	+ 140.50	221.92	0.0004	519	I AL436338 T3 end of clone BC0A0
gb_pat1:MCY78	+ 137.50	181.47	0.0715	33818	I Z77165 Mycobacterium tubercu
gb_pat1:AE001780	+ 137.00	190.88	0.0214	11997	I AE001780 Thermotoga maritima

gb_pat1:AE005497	- 135.50	189.83	0.0245	10421	I AE005497 Escherichia coli
gb_pat1:AP002562	- 135.50	157.71	1.51	270365	I AP002562 Escherichia coli
gb_pat1:AE005603	+ 135.00	187.78	0.0318	11809	I AE005603 Escherichia coli
gb_pat1:AP002566	+ 135.00	156.52	1.75	280900	I AP002566 Escherichia coli
gb_pat1:SCSP15A	+ 134.50	187.04	0.0350	11719	I X71622 Saccharomyces cerev

seq_name: gb_pat1:A84558

seq_documentation_block:

LOCUS A84558 758 bp DNA PAT 21-JAN-2000

DEFINITION Sequence 1 from Patent WO9845456.

ACCESSION A84558

VERSION A84558.1 GI:6733477

KEYWORDS Saccharomyces cerevisiae.

SOURCE Saccharomyces cerevisiae.

ORGANISM Saccharomyces cerevisiae.

REFERENCE 1 (bases 1 to 758)

AUTHORS Ebneeth M. and Sonnwald U.

TITLE 2-DEOXYGLUCOSE-6-PHOSPHATE (2-DOG-6-P) PHOSPHATASE DNA SEQUENCES AS

JOURNAL SELECTION MARKERS IN PLANTS

EBNEETH MARCUS (DE); SONNWARD UWE (DE)

FEATURES

source

1. 758

/organism="Saccharomyces cerevisiae"

/db_xref="taxon:4932"

9. 749

/note="unnamed protein product"

/codon_start=1

/protein_id="GI:6733478"

/translation="MAEFSADICLFDGCTIVSTVAEAKWTKLCYEYGVDSPLFK

HSNGATQVLRFRPKLDDTDNKGVALEKDAHSYLDTVSLIPAEMLLSIDYDT

ETQKLPERRKMTIVSGSPYLAWSFETLLKNGKRVFTGCDVANGKPRDDEGYSRA

ROLRLDLDLTKQDRLKYVFEADVPVIGAGKAMGATVGTITSSYDKSVLPDGAQVY

VCDLTQVSVAKNENGVIOVNNPLPRA"

BASE COUNT 221 a 135 c 183 g 219 t

ORIGIN

alignment_scores:

Quality: 1269.00 Length: 246

Ratio: 5.159 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-558-284-2 x A84558 ..

Align seg 1/1 to: A84558 from: 1 to: 758

1 MetLactupheseraIaAspleucysleupheaspIeuspIyThr11 17

|||||

9 ATGCGAGAAATTTCAGCGATCTATGCTTTTACCTAGCTAGTACAT 58

17 eValserThrThrValAlaAlaGluValAlaTrpThrIleuysuSTyG 34

|||||

59 AGTAGTAGACACAGTGGCGCGAGAAAGCATGAGCAAGTTGTTACG 108

34 IuTyrgIyValAspProSerGluLeuphepIeupheaspIeuspIyThr 50

|||||

109 AATACGGGTGTTGATCTTCGAGTATTTAAGCATTTCCATCGAGCAAGA 158

51 ThrIngluValIleuArgGArgPhepPheProIyLeuaspIeuspIyThr 67

|||||

159 ACACAAAGAGTTTTCAGAAAGTTTTCCTTAATTGATGATACAGCAA 208

67 pLySGIValIleuAlaIleuGluValAspIleAlaHisSerTyIleuasp 84

|||||

209 TAAAGGTTCTTCCTCTAGAAAGATATTGCCCATAGTTACTTGAGTA 258

84 hValserIleuIleuProGlyAlaGluValIleuLeuLeuSerIleuaspVal 100

```

|||||
259 CAGTAAGCCTATTCCTGTCAGAGAACTACTGTTATCGTTAGATGA 308
101 AsphrngluthrGlnLysLysLeuProGluArgLysTrpAlaIleValTh 117
|||||
309 GATACGTAGATCAAAAAAGTTACTGAAAGAAATGCCCTATCGTTAC 358
117 rSerGlySerProTyrLeuAlaPheSerTrpPheGluThrIleLeuLysA 134
|||||
359 CTCGTGTTCCCATATTTGGCATTTTCATGTTGACGACATATTTGAAAA 408
134 snValGlyLysProLysValPheIleThrGlyPheAspValLysAsnGly 150
|||||
409 ATGTGGAAAGCCCAAGTTTTCATCTACTGGGTTTGACGTGAAGACGGT 458
151 LysProAspProGluGlyTyrSerArgAlaArgAspLeuLeuArgGlnAs 167
|||||
459 AAGCCTATCCCGAGGGGTATTCAAGAGCTCGATTATTATTCGTCAGA 508
167 PleuGlnLeuThrGlyLysGlnAspLeuLysTyrValValPheGluAspA 184
|||||
509 TTTCGATTAATCTGTATGACAGAGATCTGAAGTATGTTGCTTCAGAGATG 558
184 laProValGlyIleLysAlaGlyLysAlaMetGlyAlaIleThrValGly 200
|||||
559 CACCCGCGGCGATAAAGCGCGCAAGCAATGGCGCATTAAGTGGGT 608
201 IleThrSerSerTyrAspLysSerValLeuPheAspAlaGlyAlaAspTyr 217
|||||
609 ATAAACATCCCTCGTATGACAGAGCGTTTATTGACGACAGACAGATTAA 658
217 rValValGlyAspLeuThrGlnValSerValValLysAsnAsnGluAsnG 234
659 TGTAGCTGTGATTGACACAGGTTCCGCTTAAGAACATGAAAAACG 708
234 LylleValIleGlnValAsnAsnProLeuThrArgAla 246
709 GTATTGTCATCCAGGTAAACACCTTTTGACAAAGGCC 746

seq_name: gb_pat1:A85030

seq_documentation_block:
LOCUS A85030 758 bp DNA
DEFINITION Sequence 1 from Patent EP0870836.
ACCESSION A85030
VERSION A85030.1 GI:6733776
KEYWORDS
SOURCE
ORGANISM Saccharomyces cerevisiae.
Saccharomyces cerevisiae.
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
REFERENCE
1 (bases 1 to 758)
AUTHORS Ebneth,M. and Sonnenwald,U.
TITLE 2-Deoxyglucose-6-Phosphate (2-DG-6-P) Phosphatase DNA sequences
for use as a selection marker in plants
JOURNAL Patent: EP 0870836-A 1 14-Oct-1998;
IPK GATERSLEBEN (DE)
FEATURES
source
Location/Qualifiers
1..758
/organism="Saccharomyces cerevisiae"
/db_xref="taxon:4932"
9..749
/note="unnamed protein product"
/codon_start=1
/protein_id="CAB69345.1"
/db_xref="GI:6733777"
/translation="MAFESADLCFLDGTIVSTTVAAEKAMTKLCYEVDPSELFK
HSRGARQEVLRFRPKLDPTDNKGVALLEKDIASHSLDVTSLIPGAENLLSLDVP
ETGKLLPERKMAIVTSGSPYLAISWEETILKNVGPVFLTGFDDVKNKPEPESYSA
RDLLRODLQITGKODLKYVFEADAPVGIKAGKAMGALTIVGITSYDSKSVLFDAADIV
VCDLTQVSVYKNNENGIVIOVNNPLTFA"
BASE COUNT 221 a 135 c 183 g 219 t
ORIGIN

```

```

alignment_scores:
    quality: 1269.00    length: 246
    Ratio: 5.159
Percent Similarity: 100.000    Percent Identity: 100.000

```

```

alignment_block:
US-09-558-284-2 x A85030

```

```

Align seg 1/1 to: A85030 from: 1 to: 758

```

```

1 MetAlaGluPheSerAlaAspLeuGlyLeuPheAspLeuAspGlyThrI 17
|||||
9 ATGGCAGAAATTTTACGTGATCTATGCTTTTGGACCTAGATGCTACCAT 58
17 eValSerThrThrValAlaAlaGluLysAlaTrpThrLysLeuGlyTyrG 34
59 AGTAGACACAAACAGTGGCCGACAGAAAGCATGACCAAGTTGCTTACG 108
34 IutyrGlyValAspProSerGluLeuPheLysHisSerHisGlyAlaArg 50
109 AATACGCTGTTCATCCCTCCGAGTTATTAMGCATTCTCATGCTGCAAGA 158
51 ThrGlnGluValLeuArgArgPhePheProLysLeuAspAspThrAspAs 67
159 ACACAAAGAGGTTTGGAGAGGTTTCCCTAAATTGGATGATACAGACAA 208
67 nLysGlyValLeuAlaLeuGluLysAspIleAlaHisSerTyrLeuAspT 84
209 TAAAGGTGTTCTTCCTGTAAGAAAAAGATATTGCCCATAGTTACTTGANA 258
84 hrValSerLeuIleProGlyAlaGluAsnLeuLeuSerLeuAspVal 100
259 CAGTAAACCTTATTCCTGCTGTCAGAGAACTTACTGTTATCGTTAGATGA 308
309 GATACTGACAGCTCAAAAAAGTTACTGAAAGGAATGGGCTATCGTTAC 358
101 AspThrGluThrGlnLysLysLeuProGluArgLysTrpAlaIleValTh 117
117 rSerGlySerProTyrLeuAlaPheSerTrpPheGluThrIleLeuLysA 134
359 CTCGTGTTCCCATATTTGGCATTTTCATGCTGCGAGACAAATATGAAAA 408
134 snValGlyLysProLysValPheIleThrGlyPheAspValLysAsnGly 150
409 ATGTTGGAAGCCCAAGATTTTCATTACTGGGTTTGACGTGAAGACGGT 458
151 LysProAspProGluGlyTyrSerArgAlaArgAspLeuLeuArgGlnAs 167
459 AAGCCTGATCCCGAGGGTTATTCAAGAGCTCGATTATTATTCGTCAGA 508
167 PleuGlnLeuThrGlyLysGlnAspLeuLysTyrValValPheGluAspA 184
509 TTTCGATTAATCTGTAAACAGAGATCTGAAGTATGTGCTTCGAAAGAG 558
184 laProValGlyIleLysAlaGlyLysAlaMetGlyAlaIleThrValGly 200
559 CACCCGCGGCGATAAAGCGCGCAAGCAATGGCGCATTAAGTGGGT 608
201 IleThrSerSerTyrAspLysSerValLeuPheAspAlaGlyAlaAspTyr 217
609 ATAAACATCCCTGTATGACAAAGAGCTTTTATTTGACGACGAGCAGATTA 658
217 rValValGlyAspLeuThrGlnValSerValValLysAsnAsnGluAsnG 234
659 TGTAGCTGTGATTGACACAGGTTCCGTGTTAAGAACATGAAAAACG 708
234 LylleValIleGlnValAsnAsnProLeuThrArgAla 246
709 GTATTGTCATCCAGGTAAACACCTTTTGACAAAGGCC 746

```

```

seq_name: gb_pl3:SCU03107

```

```
seq_documentation_block:
LOCUS SCU03107 1650 bp DNA PLN 15-OCT-1994
DEFINITION Saccharomyces cerevisiae 2-deoxyglucose-6-phosphate phosphatase
(DOG1) gene, complete cds.
ACCESSION U03107
VERSION U03107.1 GI:495309
KEYWORDS
SOURCE Saccharomyces cerevisiae.
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE
1 (bases 1 to 1650)
Sanz, P., Rander-Gill, F. and Prieto, J.A.
Molecular characterization of a gene that confers 2-deoxyglucose
resistance in yeast
JOURNAL Yeast 10 (9), 1195-1202 (1994)
MEDLINE 95274321
REFERENCE 2 (bases 1 to 1650)
Sanz, P.
TITLE Direct Submission
AUTHORS Submitted (03-NOV-1993) Pascual Sanz, Instituto Agroquimica y
Tecnologia Alimentos (CSIC), Cereales, Jaime Roig, 11, Valencia,
Valencia, 46010 Spain
JOURNAL Location/Qualifiers
FEATURES
source
1..1650
/organism="Saccharomyces cerevisiae"
/db_xref="taxon:4932"
/clone="10a12-13"
649..1389
/gene="DOG1"
649..1389
/gene="DOG1"
/codon_start=1
/function="phosphatase"
/evidence="experimental"
/product="2-deoxyglucose-6-phosphate phosphatase"
/protein_id="AAC48923.1"
/db_xref="GI:495310"
/translation="MAEFSADGLFDLDGTVTVAEKAMTKLCYEGVDSPELFK
HSHGARQGEVLRFPKLDPTDNKGVLALEKDLAISYLDTVSLIQAEMLILSDVDI
ETOKLPERKVAIVSGSPYLAFSMFETILKAVGKPKVITTFDVKKNCKPDEGSRSA
RDLLRODILQITGKODLKIVYFEDAVGIIKAGAMGATIVGITSIDKSVLPDAGADYV
VCDLTVSVVKNNGIVIVONNPLTRA"
BASE COUNT 501 a 280 c 333 g 536 t
ORIGIN
alignment_scores:
Quality: 1269.00 Length: 246
Ratio: 5.159 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-558-284-2 x SCU03107 ..
Align seg 1/1 to: SCU03107 from: 1 to: 1650
1 MetAlGluPheSerAlaAspLeuCysLeuPheAspLeuAspGlyThrIi 17
|||||
649 ATGGCAGAAATTTTCAGCTGATCTATGCTTTTGGACTAGATGGTACCAT 698
|||||
17 eValSerThrThrValAlaIaGluLysAlaIaTPrThrLysLeuCysTyrG 34
|||||
699 AGTGAAGACAAACAGTCGCCAGAGAAAGCATGACCAAGTGGTGTACG 748
|||||
34 LuTyrgLyValAspProSerGluLeuPheLysHisSerHisGlyAlaIaArg 50
|||||
749 AATTAAGGTGTGATCTTCGAGATTATTAAGCATTCATCGTGCAGAGA 798
|||||
51 ThGInGluValLeuAlaIaArgPhePheProLysLeuAspSPTrrAspAs 67
|||||
799 ACACAGAGAGGTTTGGACAGAGTTTTCCTAATGTGATGATACAGACAA 848
```

```
67 nLysGlyValLeuAlaLeuGluLysAspIleAlaHisSerTyrLeuAspT 84
|||||
849 TAAAGGCTGCTCTGCTCTAGAAAAAGATATTGCCCATGATTACTGGACA 898
|||||
84 hrValSerLeuIleProGlyAlaIaGluAsnLeuLeuLeuSerLeuAspAl 100
|||||
899 CAGTAAAGCCTTATTCCTGTCGACAGAACTTACTGTTATCTGTACATGTA 948
|||||
101 AspThrGluThrGlnLysLysLeuProGluArgLysTrpPalaIleValTh 117
|||||
949 GATACTGAGACTCAAAAAAGTTACTCTAAAGAGAAATGGCTATGCTTAC 998
|||||
117 rSerGlySerProTyrLeuAlaIaPheSerTrpPheGluThrIleLeuLysA 134
|||||
999 CTCTGCTTCTCCCTATTTGGCATTTTCATGTCGTGCAGACATATTGAAAA 1048
|||||
134 snValGlyLysProLysValPheIleThrGlyPheAspValLysAsnGly 150
|||||
1049 ATGTTGGAAGCCCAAGTTTTCATTACTGGCTTTGACGTGAAGAACGCT 1098
|||||
151 LysProAspProGluGlyTyrSerArgAlaArgAspLeuLeuArgGlnAs 167
|||||
1099 AAGCTGATCCCGAGGGTTATTCAAGAGCTGCTGATTATTGCGTCAAGA 1148
|||||
167 pLeuGlnLeuThrGlyLysGlnAspLeuLysTyrValValPheGluAspA 184
|||||
1149 TTTGCAATTAACTGTTAAACAGAGATCTGACATGTTGCTTTGGAAGATG 1198
|||||
184 lAprOvalGlyIleLysAlaGlyLysAlaMetGlyAlaIleThrValGly 200
|||||
1199 CACCCTGCGCATTAAGCCCGCTTAAGCAATGCGCGCAATTACTGTGCGCT 1248
|||||
201 lIleThrSerSerTyrAspLysSerValLeuPheAspAlaGlyAlaAspTy 217
|||||
1249 ATTAACATCTCTGATGATTAAGAGCGTTTATTATGACGAGAGCGAGATTA 1298
|||||
217 rValValLysAspLeuThrGlnValSerValValLysAsnAsnGluAsnG 234
|||||
1299 TGTAGTCTGATTTGACACAGGTTTCCGTGTTAAGAACAAATGAAACG 1348
|||||
234 lYlleValIleGlnValAsnAsnProLeuThrArgAla 246
|||||
1349 GTATTGTCATCCAGTAAACAACTTTGACAGAGGCC 1386
|||||
seq_name: gb_p14:YSC8179
seq_documentation_block:
LOCUS YSC8179 4413 bp DNA PLN 04-SEP-1997
DEFINITION Saccharomyces cerevisiae chromosome VIII cosmid 8179.
ACCESSION U00062 U00093
VERSION U00062.1 GI:488162
KEYWORDS
SOURCE Saccharomyces cerevisiae strain=5288C (AB972).
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE
1 (bases 1 to 4413)
Johnston, M., Andrews, S., Brinkman, R., Cooper, J., Ding, H., Doyer, J.,
Du, Z., Favell, O.A., Fulton, L., Gattung, S., Geisel, C., Kirsten, J.,
Kucaba, T., Hillier, L., Jier, M., Johnston, L., Keppeler, D.,
Langston, Y., Latreille, P., Louis, E., Macri, C., Mardis, E.,
Mouster, L., Nhan, M., Rifken, L., Riles, D., St. Peter, H., Thornton, L.,
Trevisan, E., Vaudin, M., Vaughan, K., Vignati, D., Wilcox, L.,
Willis, A., Wilson, R., Wohlman, P. and Waterston, R.
Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
VIII
JOURNAL Science 265 (5181), 2077-2082 (1994)
MEDLINE 94378003
PUBMED 8091229
REFERENCE 2 (bases 1 to 4413)
AUTHORS Du, Z.
TITLE The sequence of S. cerevisiae cosmid 8179
```

JOURNAL Unpublished (1994)
 REFERENCE 3 (bases 1 to 44113)
 AUTHORS Waterston, R.
 TITLE Direct Submission
 JOURNAL Submitted (19-MAY-1994)
 REFERENCE 4 (bases 1 to 44113)
 AUTHORS Jia, Y. and Cherry, J. M.
 TITLE Direct Submission
 JOURNAL Submitted (04-SEP-1997) Department of Genetics, Stanford University, Saccharomyces Genome Database, Stanford, CA 94305-5120, USA

Curated by:
 Saccharomyces Genome Database
 URL: <http://genome-www.stanford.edu/>
 e-mail: yeast-curator@genome.stanford.edu

Submitted by:
 Genome Sequencing Center
 Department of Genetics, Washington University,
 St. Louis, MO 63110, USA
 e-mail: mj@sequencer.wustl.edu

NEIGHBORING COSMID INFORMATION:

This sequence includes nucleotides 1-43838 of cosmid 8179 plus 75 bp of cosmid 9181 (cosmid adjacent to the right end of 8179). This sequence overlaps with cosmid YSC88025 on the right by 200 bp. The adjacent cosmid to the left is YSC88082.

Location/Qualifiers
 1. 44113
 /organism="Saccharomyces cerevisiae"
 /db_xref="taxon:4932"
 /map="YII1"
 /strain="S288C (AB972)"
 /complement(539. 1993)
 /gene="SLR2"
 /complement(539. 1993)
 /gene="SLR2"
 /note="YHR030C: Other names MPK1, SLR2, BYC2."
 /codon_start=1
 /product="Slr2p: Putative Ser/Thr protein kinases"
 /protein_id="AAB68912.1"
 /db_xref="GI:488177"
 /translation="MAKDIERHFKVFNODPSVYKRPQLIKIGHGAYGYCSARFAEAAETTYAIKVTWVESKTLCKRSLEKLIRFRGKNITTCYDMDIVYPPGSGINGLYIEELMECDHQILIKSGPLTDHAYQSFYQILGKLYHSADYVLRDLKQGNLVMACQQLKIDCFGLAGYSENPVNSQFLTEYVATRWRAEIMLSYGYSYTKADVMSAGCILAIEFLGKPIFKGKYNOJLLOLVGTPDETLRIGSKNODYTHQLGFIPIKVPVNLYPNANSQALDLBOMLAEDPQKRTVDELEHPYLSIMHPADBPVCSSEKFEFSFESYDMDEDKOMVIOYQDFRFLYRQPLFEORQLOLQOQOQOQOQOQOQOQSDVDNGMAAASEENYPKOMATSNVAPQOESFGIHSQVLPKADPFPPEDESMEMKRPATGNTADIPQNDNGTLDDLEKELFGLDKRP"

gene
 CDS
 complement(2448. 4619)
 /gene="YHR031C"
 /complement(2448. 4619)
 /gene="YHR031C"
 /note="Similar to Pif1p (S. cerevisiae), mitochondrial DNA repair and recombination protein."
 /codon_start=1
 /evidence="not_experimental"
 /product="Yhr031cp"
 /protein_id="AAB68913.1"
 /db_xref="GI:488178"
 /translation="MFRSHASGNKKQWSSKSSNGSTPAASAGSGHAYRQQTLLSFEFGCCKSAASAKNSTIILDESDEGNRMTAPPRRLIRNNSSSLFESQSGFGDDDPAAERKTLVDPRNLNYSKSSLSMTSLHTASATQKTYHDEDETLREYTVSKNSROLSESTINIEDSSMKLSTDSERPAKRSKSPMEFGKLTVPKTIKIPILKRTVSNMDSNNHRASSPVVLTMEQERYVNLIVKRTNVEYTGSGAGCGYVLIOTIIRQSLSYGKESATATAGLAAYTIGGSTLHKSSGIGKNTIDQVKKIQSQKDLAARWTKVLIIDETISWDGNLQKLEQIARIRKNDPFGQIQLVLTGDFQLPPAKKDEHNVAFCFSEMWKRIQKTIILTFKVRQDQNLILNAIRYGEITVDIAKTIIRNLNDIDYA

gene
 CDS
 complement(8614. 9648)
 /gene="YHR034C"
 /complement(8614. 9648)
 /gene="YHR034C"
 /codon_start=1
 /evidence="not_experimental"
 /product="Yhr034cp"
 /protein_id="AAB68914.1"
 /db_xref="GI:488179"
 /translation="MADFLRPYKORHRENEKYVDAADGSYKIEPIADFYIKTKLLSANGPEKLDGRKVFVNWCHSLVPKREVDNRIYEPPLIIONEMELPITTSYRMDHDKKGOEYVWDCINSDCSRWICDDIOLRLIIVEMCESECIKSDSVLCNDRIAPFKMKKGAELPAALEVINDLHODYKAKKMIIEEGGDPMSIICRGRDDDDNDPDDGTLPPLPTENKISGAKIEIDKNTAHRKLCAPAPAPAPHCQOQDVEPEYKMKRFPGAAYKRLIIEKAKNSKDRSPSYNRAENLIYINGKLSITLPDPIYVNNADITFIHRRKRTIYIT"

gene
 CDS
 complement(8614. 9648)
 /gene="YHR035W"
 /complement(8614. 9648)
 /gene="YHR035W"
 /note="Weak similarity to Sec23p (S. cerevisiae)."
 /codon_start=1
 /evidence="not_experimental"
 /product="Yhr035wp"
 /protein_id="AAB68909.1"
 /db_xref="GI:488174"
 /translation="MCSPTNPLYPFSSDAYTOWYDQNLKTKCGAYSNACSLREONVWTLFCQNSNAELPLVPSNTTTLTSAKEILSKRTIMLIDICDHEINLYVSLICNNYTRQOEPLSLITIQSGVILHNAVNRRAVPSINEFMKRYNLDKINASYAEKKSIEINQESYWFDSKSLRKLRLREICKIANKNVISKRDKCTGLALFEVSVLASQ

DGIAPTELYATREVELSNVKKLOSLPGDLYEFAKAVDANAPERYOAILDSSLMVEKVALKEDAOVMMMLKNKPDVELVNGSLGKVFEEVNSHILVKKKELIKIYDDVDMDRIVSRVIGNPDLKESKEPFIIDIPRENVGLERTQIPMLCNALSIHKQGOITQRLKDLRIFEBAGVYVALSRVMTDTLQVLMFDPGKIRTNERNVDFYKRIETLK

gene
 CDS
 complement(8614. 9648)
 /gene="YHR032W"
 /complement(8614. 9648)
 /gene="YHR032W"
 /codon_start=1
 /evidence="not_experimental"
 /product="Yhr032wp"
 /protein_id="AAB68911.1"
 /db_xref="GI:488176"
 /translation="MSKQSHTTNDRSSIIYSTSVGAGLFTPADYIPQSENLJIEGEOBGSSEEPSTYGNDETEREGEYHSLDANNSRTIQQAMQOQYSHRKRILDEERDLITDNKLISOHONGGDIESHQGAIGDEERAEALIANWESIESGQISTTFKRETOVITNNALPLIFPIILQNSLSIASIFSVSHTGKELGVTIGSMATITGLAAIQGLCTCUDTCAQAYGAKNYHLVGIVQRCVAVITILAFPMYVMEVMSKILALMIPEBELCALANAYLRVATFAGVPGEILFECGRFLQOQGFHASTIYLVCPADLNALMYLLVMNDKIGTIGLAPLSVYINVMITGLIYAMTTHKREPKCWNGIIPKROAFKMKRKMINAIPGVVVAEAEFLGPEVLTIFASHLGTDLGAGSIYATTISLAQVPEFSVSTSTRVANTGASLYDSQMITCRVSLTSPVCSMMMFVICTKEQIASLFTESAIVKMWVDPLPLAFMQLPDAPNASTAGLRQGRKQKRWVHQPSKILLPRCAHGICVSI

gene
 CDS
 complement(8614. 9648)
 /gene="YHR033W"
 /complement(8614. 9648)
 /gene="YHR033W"
 /note="Highly similar to Pro1p (Gamma-glutamyl kinase)."
 /codon_start=1
 /evidence="not_experimental"
 /product="Yhr033wp"
 /protein_id="AAB68910.1"
 /db_xref="GI:488175"
 /translation="MTKAVTIVIKGSSSLVDESTEKPKISIMTLIVETYNLRMGHKVLIIVSSGIAVGLDALNIPIKPKQISEVQAIAVAGQRLIARNNMLFSQGEQTAQILTRNDILRMQYNNARNTINELAMGYIPIVENDUTLISEIEFGDDTLATAALVGADPEFLITDVCGLTYNDPRTNDARPILVPELSEGLPGVNTSSGSGSVGGAMRTKLIADLASNAGETIYMKSDREYVRYKYIDYTOHRRPRRHGNGCQOQFLODTLELDQRTDVPMTHTFLANDNKKHLKNEPILHGLITKGAIIIDENSYTDLKSDMAELTPNNAVLEVNDNFHECEVDLKGIRLNGEILDISKPIQSGVRSNYSYTLLELAKIKGLPSEKIHVDYGVSEVYAHRENIAFPQP"

CSLSAVCHIVFLNGPCTKGGKVMRSREGESMRONHFEESKSQOLSKSPTKFYK
 MEKFRANOSLIYEFLIASLDQIGILEMSPLITSSMAVSOPDSFNDERFAMSPQYINL
 RDHNAIYNCHSKIMTAKNAIVKDEPKYSLNKNSPLPLEISLGHNSAPLOPOTF
 ENOTKRYRIETELLIPKANRSFGAONEIVFSKKIJAISRIIDSPAYSSKTKELMOLF
 LLPNOIRKGDVNMVNLICMCHYHITSPIISVNTSPDEYETLELHRIANSKDTCLSLC
 KPFISYSDLKHDMIVLVDPIITRAOILDDDKTICVDGGSYLVLRGKLLKEGREL
 CKLANDIOPFLYETKGTGSDRFLKSIIPLDIDKETLGTEDMTFENYFNLT
 DLSGSK"

gene
 11994..13409
 /gene="YHR036w"
 11994..13409
 /gene="YHR036w"
 /codon_start=1
 /evidence="not_experimental"
 /product="Yhr036wP"
 /protein_id="AAB68908.1"
 /db_xref="GI:488173"

/translation="MSEFENLSIRDSFTSGMEHVEDELGLSDLSISKQGPPLSPOLI
 NRPMPHPSSPLRNTLDESAKADEEEDMEIDEVDITSEEEYNEPIETHTEA
 TENAVEEIEATPEEROKOEKNSODSVEEVENIVSPHRSVIFALLSPDLGYAAA
 TKVEGVPLPSPANODNESSNNNAEGDIINEVEDEIKSLGHNKSIDEPYDNEAFKL
 EIKRELRSKSYOPLOVSFNTHNYEYSPDKDITKYSILTKPWHNKIDEPYDNEAFKL
 PKPSPNHPASRASAYALMSYOLFPLNATTTVTISFISLIALOKDKLWBEORKH
 ELQYSRIGOEYLTNRCONPTGLPALGEOCAIMKOCMDRNDIDFRASDTLAKLFG
 DIINSFIDPLNKKTLFVIFCGVITWCFSSNFLGFRVAKSYGNGIKYPLPSSPKSP
 TSETHSSMTASCEDSHLKQ"

gene
 13626..15353
 /gene="PUT2"
 13626..15353

CDS
 13626..15353

alignment_scores:
 Quality: 1269.00 Length: 245
 Ratio: 5.159 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-558-284-2 x YSCH8179/rev ..

Align seg 1/1 to reverse of: YSCH8179 from: 1 to: 4413

1 MetAlaGluuPheSerAlaAspLeuGlyLeuPheAspLeuGlyThr11 17
 26457 ATGCGAAGATTTCAGCTGATGCTATGCTTTTGACCTAGATGCTACCAT 26408
 17 eValSerThrThrValAlaIaGluLysAlaTrpThrLysLeuGlyTyrG 34
 26407 AGTGAGACACAGTGGCCGACAGAAAGCATGACCAAGTTGCTTACG 26358
 34 LuTrpGlyValAspProSerGluLeuPheLysHisSerHisGlyAlaArg 50
 26357 AATACGCTGTTCCTTCCTCCAGCTTATTAAAGCATCTCATGTGCAAGA 26308
 51 ThrGluGluValLeuArgArgPhePheProLysLeuAspSerThrAspAs 67
 26307 ACACAGAGGCTTTTGAGAAAGCTTTTCCCTAAATTCGATGATACAGCAA 26258
 67 nLysGlyValLeuAlaLeuGluLysAspIleAlaHisSerTyrLeuAspT 84
 26257 TAAAGCTGTTCTGCTCTAGAAAAAGATATTGCCCATAGTACTTGACA 26208
 84 hValSerLeuIleProGlyAlaGluAsnLeuLeuSerLeuAspVal 100
 26207 CAGTAGACCTTATTCTGCTGACAGAACTTACTGTTATCTTAGATGTA 26158
 101 AspThrGluThrGlnLysLysLeuProGluArgLysTrpAlaIleValTh 117
 26157 GATACGACAGCTCAAAAAAGTACCGAAAGCAAAAGCGCTATCGTTAC 26108
 117 rSerGlySerProTyrLeuAlaPheSerTrpPheGluThrIleLeuLysA 134
 26107 CTCTGGTTCCTCATATTGCGCATTTTTCATGCTTCGACACATATTGAAA 26058
 134 snValGlyLysProLysValPheIleThrGlyPheAspValLysAsnGly 150

|||||
 26057 ATGTGGAAAGCCAAAGTTTTCATTACTGGGTGAGCTGAAGAAGCGT 26008
 151 LysProAspProGluGluGlyTyrSerArgAlaArgAspLeuLeuArgGlnAs 167
 26007 AAGCTGATCCCGAGGCTTATTCAGAGAGCTCGTATTTATTCGCGTCAAGA 25958
 167 pLeuGlnLeuThrGlyLysGlnAspLeuLysTyrValValPheGluAspA 184
 25957 TTTCGAATTAACTGTTAAACAGAGATCTGAATATGTTCTTCCGAAGATG 25908
 184 lAprOvalGlyIleLysAlaGlyLysAlaMetGlyAlaIleThrValGly 200
 25907 CACCCGTGGCATTAAGCGCCGCAAGCAATGGCGCATTAAGTGTGGCT 25858
 201 lLeThrSerSerTyrAspLysSerValLeuPheAspAlaGlyAlaAspTy 217
 25857 AATACATCTCTGTTATGACAAAGCGTTTATTTTACGACGAGACAGATTA 25808
 217 rValValCysAspLeuThrGlnValSerValValLysAsnGluAsnG 234
 25807 TCTAGTCTGTCATTGACACAGGTTTCGCTGTTAAGACATGAAACG 25758
 234 lYlIleValIleGlnValAlaAsnProLeuThrArgAla 246
 25757 GTATTGTCATCAGGTAAACACCTTTGACACAGGCGC 25720

seq_name: gb_stsl:CNS0616P

seq_documentation_block:

LOCUS CNS0616P 826 bp DNA STS 10-JAN-2001
 DEFINITION T3 end of clone AS0A018C02 of library AS0A from strain CLIB 533
 of Saccharomyces bayanus, sequence tagged site.

ACCESSION AL399911
 VERSION AL399911.1 GI:12155045

KEYWORDS STS.

SOURCE Saccharomyces bayanus.

ORGANISM Saccharomyces bayanus

REFERENCE 1 (bases 1 to 826)
 Bon,E., Neuveglise,C., Casaregola,S., Artiguenave,F., Wincker,P.,
 Aigle,M. and Durans,P.
 Genomic Exploration of the Hemiascomycetous Yeasts: 5.
 Saccharomyces bayanus var. uvarum
 FEBS Lett. 487 (1), 37-41 (2000)

JOURNAL

PUBMED

AUTHORS

REFERENCE

AUTHORS

JOURNAL

PUBMED

AUTHORS

REFERENCE

AUTHORS

JOURNAL

PUBMED

AUTHORS

REFERENCE

AUTHORS

JOURNAL

PUBMED

AUTHORS

REFERENCE

AUTHORS

JOURNAL

PUBMED

AUTHORS

REFERENCE

AUTHORS

JOURNAL

PUBMED

AUTHORS

REFERENCE

AUTHORS

JOURNAL

PUBMED

2 (bases 1 to 826)
 Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
 Bolotin-Fukuhara,M., Bon,E., Brotier,P., Casaregola,S.,
 de-Montigny,J., Dujon,B., Durans,P., Lepingle,A., Llorente,B.,
 Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
 Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
 Wincker,P. and Weissenbach,J.
 Genomic Exploration of the Hemiascomycetous Yeasts: 1. A set of
 yeast species for molecular evolution studies(1)
 FEBS Lett. 487 (1), 3-12 (2000)

3 (bases 1 to 826)
 Genoscope.
 Direct Submission
 Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
 2 rue Gaston Cremieux, CP 5706, 91057 Evry cedex, FRANCE. (E-mail :
 seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)

This STS is part of a random genomic sequencing program of thirteen
 yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
 exiguus, Saccharomyces servazii, Zygosaccharomyces rouxii,
 Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
 lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
 angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
 Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
 5 kb were prepared and both extremities were sequenced. See
 keywords for description of this sequence and for the sequence of

137 yspProlyValIphelIerhrglyPheaspValIysasnGlyIysProasp 153
 388 AGCCAAAGGTTTCACTTACTGCAAAATGATGTTCCGAAAGTAAAGCCGAAT 437
 154 ProgluGlyTyrSerArgAlaIarGAsPleuLeuArgGlnAspLeuGlnLe 170
 438 CCCGAAGC.TACTTGAGGACGATTTAGTCAATTAAGAACAACAACAACT 486
 170 uThrlGlySglnAspLeuLysTyrValValAlpheGluAspAlaProValG 187
 487 AGTTGCT...CATGAATTTTCGCTGCTGTTTCGAAAGACCCCGGTGG 533
 187 IyIleLysAlaGlyLysAlaMetGlyAlaIleThrValaGlyIleThrSer 203
 534 GTATTCAAGCCGGTATTGCTGCAGAGGTTTCATGTTATTTGGATTGCAACT 583
 204 SerTyrAspLysSerValLeuPheAspAlaGlyAlaAspTyrValValGly 220
 584 ACTTTGATTAAGGATTAATTAAACGAGATCATCATTTGTTAGTGA 633
 220 SASpleuThrglnValSerValIallySAsnAsnGlnGlyIleValI 237
 634 AGATTGTCAAAATACACATTACGAAACTACTTCACATCATTCACAT 683
 237 IeGlnValAsn 240
 684 TAGATTGGAT 694

seq_name: gb-pl3:CA20C1

seq_documentation_block:

LOCUS CAC20C1 37968 bp DNA PLN 05-NOV-1998
 DEFINITION C.albicans cosmid Ca20C1.
 ACCESSION AF033391
 VERSION AF033391.1 GI:3850121

KEYWORDS 2-deoxyglucose-6-phosphate phosphatase; DEAD box helicase; DNA repair protein; Dog2; ER retention signal; membrane protein; methyltransferase; pyridoxine kinase; ribosomal protein L32e; RING finger protein; serine/threonine protein kinase; transfer-RNA Ala; transport protein; zinc finger protein.
 Candida albicans.

SOURCE

ORGANISM Candida albicans
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; mitosporic Saccharomycetales; Candida.

REFERENCE 1 (bases 1 to 37968)
 Tait,E., Simon,M.C., King,S., Brown,A.J., Gow,N.A. and Shaw,D.J.
 A Candida albicans genome project: cosmid contigs, physical mapping, and gene isolation
 Fungal Genet. Biol. 21 (3), 308-314 (1997)

JOURNAL 97435544
 MEDLINE Article No. FG970983
 REMARK 2 (bases 1 to 37968)
 AUTHORS Oliver,K. and Harris,D.
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 37968)
 AUTHORS Bartell,B.G. and Rajandream,M.A.
 TITLE Direct Submission
 JOURNAL Submitted (04-NOV-1998) Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: bartell@sanger.ac.uk
 Cosmids supplied by Prof. Duncan Shaw, [3] Department of Molecular and Cell Biology, The Institute of Medical Science, University of Aberdeen, Foresterhill, Aberdeen, AB25 2ZD, UK

COMMENT

Notes:
 Funding: sequencing funded by Beowulf Genomics Ltd. CDS are numbered using the following system eg CAC20C1.01c. CA (C. albicans), C20C1 (cosmid name),
 .01 (first CDS), c (complementary strand).
 The more significant matches with motifs in the PROSITE/PFAM databases are also included but some of these may be fortuitous.
 IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a

FEATURES
 source small overlap between neighbouring submissions. The map position of cosmid Ca20C1 is unknown.
 Location/Qualifiers
 1..37968

/organism="Candida albicans"
 /strain="1161"
 /db_xref="taxon:5476"
 /clone="cosmid Ca20C1"
 96..102
 /note="possible tata box, tataaaa upstream of Ca20C1.01"
 172..894
 /gene="Dog2"
 /note="Ca20C1.01"
 172..894
 /gene="Dog2"
 /note="Ca20C1.01"
 172..894
 /gene="Dog2"
 /note="Ca20C1.01, Dog2, len: 240 aa, similar to a family of proteins that include S. cerevisiae 2-deoxyglucose-6-phosphate phosphatase 2 (EC 3.1.3.-) (2-DOG-6-P 2), DOG2 OR YHR043C, DOG2_YEAST P38773, fasta scores opt: 476, (43.0% identity in 230 aa overlap) and DOG1_YEAST P38774, 2-deoxyglucose-6-phosphate phosphatase 1 (EC 3.1.3.-) (2-DOG-6-P 1). DOG1 OR YHR044C (246 aa) opt: 462, (41.4% identity in 239 aa overlap)"
 /codon_start=1
 /transl_table=12
 /product="putative 2-deoxyglucose-6-phosphate phosphatase"
 /protein_id="CA21923.1"
 /db_xref="GI:3850122"
 /db_xref="GI:3850123"

gene
 CDS
 1136..2860
 /gene="Ca20C1.02"
 1136..2860
 /gene="Ca20C1.02"
 /note="Ca20C1.02, len: 574 aa, similar to RNA helicases eg S. cerevisiae DBP9_YEAST DBP9. (594 aa), fasta scores, opt: 1568, E(): 0. (62.6% identity in 588 aa overlap); contains PS00017 ATP/GTP-binding site motif A, pfam match to entry DEAD PF00270, DEAD and DEAH box helicases conserved C-terminal domain"
 /codon_start=1
 /transl_table=12
 /product="dead box helicase"
 /protein_id="CA21924.1"
 /db_xref="GI:3850123"

misc_feature
 misc_feature
 misc_feature
 gene
 1136..1903
 /gene="Ca20C1.02"
 /note="pfam match to entry DEAD PF00270, DEAD and DEAH box helicases, score 220.48"
 1310..1333
 /gene="Ca20C1.02"
 /note="PS00017 ATP/GTP-binding site motif A"
 1961..2335
 /gene="Ca20C1.02"
 /note="pfam match to entry helicase C PF00271, Helicases conserved C-terminal domain, score 44.51"
 complement(2991..3620)
 /gene="Ca20C1.03c"


```
CDS      complement(2991..3620)
        /gene="Ca20C1.03c"
        /note="Ca20C1.03c, unknown len: 228 aa, slight similarity to 014142, hypothetical 23.4 kDa protein, (206 aa), fasta scores, opt: 207, E(): 2.6e-06, (26.6% identity in 177 aa overlap)"
        /codon_start=1
        /transl_table=12
        /product="hypothetical protein"
        /protein_id="CAA21925.1"
        /db_xref="GI:3850124"
        /translation="MSLPFAEKVDADLEKSMASDKNYHLHPTSLTVKKLLFALLIESTISCEFYITDRIGLHPMLAPAMLCESGALAOASNQYRKRFNLKIFRWNGFINGFYSAMISLVTRVDFNLYFRILDDSCGTMPQLFNINLAIDGGSEVSSATRIAFFSLSKYSCFMFPFSIVSFIPETLMPCNCANLFENLIISKIA"
        /length=5455
        /gene="Ca20C1.04"
        /length=5455
        /note="Ca20C1.04"
        /note="Ca20C1.04, unknown membrane protein, len: 510 aa, similar to eg YHK4_YEAST P38776 hypothetical 57.8 kDa protein in APL-SMF2 intergenic region YHR048W. (514 aa), fasta scores, opt: 1139, (47.4% identity in 468 aa overlap)"
        /codon_start=1
        /transl_table=12
        /product="hypothetical membrane protein"
        /protein_id="CAA21926.1"
        /db_xref="GI:3850125"
        /translation="MEIEFEKKNDLLEDFTTNRVYLEPELLEKTSDEPLYIFDGSDDKEDASLRSHKMTVIIGLISGVCTLTSSVALATPDIIMHDLSHEVALNTSFDLELDGGCGFLISEFHGRIRIVICGLTMTEPFELTAIFSNICAMVEGRMSGCLGSSSVSYAASFADLFKARKCKGDANKELSLAILIYSVAPFLGPSGLISCFINTHLWYMTGYVFYLMSGVELLAVLVLPETHPEPVRLORAKRIKRTGDNRYVAADKO TTSLYESVLTSSKRPIILLIKDMNFALLNYTGTLVVYVLFPAFPISTSYHPTII SEQSFIAGLVGWNVSTGISPFITDKCYVALLRKNNGVAKPEPRIPDLWGAEPVPI GLFPIIAYSTYSHIHMIAPIGSAVYGSCFTLVFNSIPAYIVEALVAASAMAINSLT RVSYSICFPLRCFLQMYKGLGIQMATTLLAFACLMIPILFFKYGYGLRSRSPYAWS VD"
misc_feature    6255..6261
                /note="potential tata box, tatlaaac, upstream of Ca20C1.05"
                /join(6298..6323,6407..6836)
                /gene="Ca20C1.05"
                /join(6298..6323,6407..6836)
                /gene="Ca20C1.05"
                /note="Ca20C1.05, unknown, spliced, len: 151 aa, similar to S. cerevisiae YN21_YEAST P53834, hypothetical 17.2 kd protein in MRPL10-ERG24 intergenic region, YNL281W or N0589, fasta scores, YN21_YEAST P53834 Opt: 384, (39.2% identity in 153 aa overlap)"
                /codon_start=1
                /transl_table=12
                /product="conserved hypothetical protein"
                /protein_id="CAA21927.1"
                /db_xref="GI:3850126"
                /translation="MVVNPNMNMHMIDKNCILPMKSDYIKENIIDTTEDSDPRVYVA VDSVGCDGMQRKXVLCITMDLRLOSGAISAIKMGEEFETSATIVPEYHHOD KDEYFEIESLSQSKSEIKRYVPILKEKLRFDPDLLAHAKDVHATD"
                /length=6324
                /note="6329"
                /note="splice donor sequence, gtagtg, between exons 1 and 2 of Ca20C1.05"
                /length=6387
                /length=6406
                /note="splice branch and acceptor sequence TACTACAGdattaatatag between exons 1 and 2 of Ca20C1.05"
                /complement(6909..9200)
                /gene="Ca20C1.06c"
                /complement(6909..9200)
                /gene="Ca20C1.06c"
                /note="Ca20C1.06c, unknown, len: 763 aa, similar to S. cerevisiae Q03388, D8035.2BP (810 aa), fasta scores, opt: 339, E(): 5.4e-29, (28.3% identity in 842 aa overlap)"
                /codon_start=1
```

```

/transl_table=12
/product="conserved hypothetical protein"
/protein_id="CAA21928.1"
/db_xref="GI:3850127"
/translation="MSDSEEEVFESLIATRRSRNAGLRKQLIALEQDSNELAATT
OFLTEDDENVNLFDQEDGEDEFFIDEVDNANLNGIDEEGTEKMENTNRKRKHTSS
DGESEEDENDTSNINMADPEILSDSELSLTDSDSEGEKLEKQELKKRKLKQSLIPT
IKQSVESATLKKKKOKRPTLVATDSLILANRRSSRAAIVESKALVDKLEKARARA
KYVHERKQOVELQBERLAEVATEKANESLRLEQELVKREKOROLLKSRILK
HNIVRLSSETFVPIIEVDNARROYDYMKIKKKLKKQNLAEKMDPEPTRPFTI
DYDPYORKILAEKKAOEONEIKELIESGSKKNDNDENIOLVPTST
NMDNNNNINTECFIKVEDSTOEIKAEETNKIDSSSKNTNISEAPSPETESTKOE
KISDTENESESENTSOTEBNLTNDNSEVKKENEEDTNNKILETPVDNKKPDQNLIT

```

[illegible]

```

3112 GATCGGAGTATGTCGTCCTTGAACATGACCCGGGGCATTAAGGCCGG 3063
191 ylysaIameGtGAlaIaIleThrValGlyIleThrSerSerTyrAspLys 208
|||||
3062 CAAACGATATGGCGGCATTTACTGTGGGTATTAACATCTCTGTATGATTA 3013
208 eValleuPheAspIaGAlaIaAspTyrValValCysAspLeuHrgin 224
|||||
3012 GCGTTTATTGTAGCGCAGGTGCAGATTATGTGCTGTGATTTGACACAG 2963
225 ValSerValValIlyAsnAsnGluAsnGlyIleValIleGlnValAsnAs 241
|||||
2962 GTTTCGGTGTTAAGACATGAGAACGGTATCGTTATTCACGTTAAACA 2913
2912 CCTTTGACGAGA 2900
241 nProLeuThrArg 245
|||||
seq_name: gb_sts1:CMS07CNW

seq_documentation_block:
LOCUS CMS07CNW 1017 bp DNA STS 12-JAN-2001
DEFINITION T7 end of clone BD0AA004D08 of library BD0A from strain CBS 94 of
Candida tropicalis, sequence tagged site.
ACCESSION AL439410
VERSION AL439410.1 GI:12222823
KEYWORDS STS.
SOURCE Candida tropicalis.
ORGANISM Candida tropicalis.
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE 1 (bases 1 to 1017)
Blandin,G., Ozler-Kalogeropoulos,O., Wincker,P., Artiguenave,F. and
Dujon,B.
Genomic Exploration of the Hemiascomycetous Yeasts: 16. Candida
tropicalis
FEMS Lett. 487 (1), 91-94 (2000)
2 (bases 1 to 1017)
Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
Boletín-Pukhara,M., Bon,E., Brothier,P., Casaregola,S.,
de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Lorente,B.,
Malpertuy,A., Neuvéglise,C., Ozler-Kalogeropoulos,O., Potier,S.,
Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
Wincker,P. and Weissenbach,J.
Genomic Exploration of the Hemiascomycetous Yeasts: 1. A set of
yeast species for molecular evolution studies(1)
FEMS Lett. 487 (1), 3-12 (2000)
3 (bases 1 to 1017)
Genoscope.
Direct Submission
Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
seqlife@genoscope.cns.fr - Web :
http://www.genoscope.cns.fr)
This STS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.

FEATURES
Source
Location/Qualifiers
1..1017
/organism="Candida tropicalis"
/strain="CBS 94"
/db_xref="taxon:5482"
/clone="BD0AA004D08"
/clone_lib="BD0AA"
/note="end : T7"

```

```

misc_feature complement(<346..>1014)
/Note="similar to Saccharomyces cerevisiae ORF YER062c (
HOR2 : Di-glycerol phosphatase 1
similar to Saccharomyces cerevisiae ORF YII053w ( RHR2 :
Di-glycerol phosphatase 1"
/evidence=no_experimental

BASE COUNT      350 a      189 c      179 g      297 t      2 others
ORIGIN

alignment_scores:
    Quality: 333.50      Length: 198
    Ratio: 2.508      Gaps: 4
Percent Similarity: 67.172      Percent Identity: 39.899

alignment_block:
US-09-558-284-2 x CNS07CWN/rev ..

Align seg 1/1 to reverse of: CNS07CWN from: 1 to: 1017

37 Valasprosergiuleupheliyserrhisserhsglyalaargthrgl 53
|||||.....:|||||.....:|||||.....:|||||.....:
963 GTTATCTCTGAGAATATTAGAACCTCCATGCTGCTACCTACGA 914
|||||.....:|||||.....:|||||.....:|||||.....:
53 uvalleuargargpherpheprolysleuaspsphraspsnlysglyv 70
|||||.....:|||||.....:|||||.....:|||||.....:
913 TGTATGCGCCAAATGCTCCAGCA.....GATCGTATGTCGACAAAG 870
|||||.....:|||||.....:|||||.....:|||||.....:
70 alleualaileuglulysaspillealahiserryrleuaspthvalser 86
|||||.....:|||||.....:|||||.....:|||||.....:
869 TTACTGCTGGGAAGTCCATTCAGATCTTTGTGTAATACCCAAA 820
|||||.....:|||||.....:|||||.....:|||||.....:
87 leuileproglialaiaulasnleuleuserleuaspsvalaspthrgl 103
|||||.....:|||||.....:|||||.....:|||||.....:
819 CCAATTCGGGTGCTGCTGTAATGCTTAAATCTTTGATTAATTCCTCA 770
|||||.....:|||||.....:|||||.....:|||||.....:
103 uthrghllylsleuprogluarglystrpalailevalthrserglys 120
|||||.....:|||||.....:|||||.....:|||||.....:
769 AGAACAAACGAAACGCTCAACAAAGATGGCTGCTGCTGCTGCTGCTA 720
|||||.....:|||||.....:|||||.....:|||||.....:
120 erprotyrleualaphesertrrphnegluthrileulelysasnvalgly 136
|||||.....:|||||.....:|||||.....:|||||.....:
719 CTTTACCATTTAGCCACCAATGCTTGAAA.....TTGTTAACCAFTGAA 676
|||||.....:|||||.....:|||||.....:|||||.....:
137 lysprolysvalpheilethrghllypheaspyllylsasnlylylproas 153
|||||.....:|||||.....:|||||.....:|||||.....:
675 AAACGACATTTTTCATCTACTGCCGAAAGGTCACCTAAAGGTAACCCCA 626
|||||.....:|||||.....:|||||.....:|||||.....:
153 pprogluglytyrserargalaargaspheuleuarglinaspheuglnl 170
|||||.....:|||||.....:|||||.....:|||||.....:
625 TCCACAAAGGTTACCAAGCGCTAGAGATCTTTG.....592
|||||.....:|||||.....:|||||.....:|||||.....:
170 eutnrglylysglnaspheulystyr.....Valvalphelguaspala 184
|||||.....:|||||.....:|||||.....:|||||.....:
591 .....GGTTACACAGATGCTCAGCAAAAGTGGCTTTTCGACAGATGCT 547
|||||.....:|||||.....:|||||.....:|||||.....:
185 provalaiailelysalaiaiametgylalaiaietnvalaiaia 201
|||||.....:|||||.....:|||||.....:|||||.....:
546 CCAAGTGTATCACTGCTGCTGTAAGAGTCTGCGCATTTGCATTTGGTAT 497
|||||.....:|||||.....:|||||.....:|||||.....:
201 ethtserseertyraspylservalleupheaspylalaiaaspyrv 218
|||||.....:|||||.....:|||||.....:|||||.....:
496 CTGTTTACTATATGACCAAGAAAGTTAGAAATCTGCTGCCAACATTTG 447
|||||.....:|||||.....:|||||.....:|||||.....:
218 a1vala1cysaspheuthrghlnvalservalvalylsasnasnclu 232
|||||.....:|||||.....:|||||.....:|||||.....:
446 TCGCGCAAGACTTGTCAAGCTTCAAGATTTGAATCTTACACAAA 403
|||||.....:|||||.....:|||||.....:|||||.....:

seq_name: gb_p14:YSPRHR2C
seq_documentation_block: 925 bp mRNA PLN 10-FEB-1999
LOCUS YSPRHR2C
DEFINITION Yeast Rhr2 mRNA, complete cds.

```

```

ACCESSION D50471
VERSION D50471.1 GI:1236749
KEYWORDS Rhr2.
SOURCE Saccharomyces cerevisiae (strain:RS16) cDNA to mRNA.
ORGANISM Saccharomyces cerevisiae
REFERENCE Enkariota: Fungi: Ascomycota: Saccharomycotina: Saccharomycetes:
AUTHORS Saccharomycetales: Saccharomycetaceae: Saccharomyces.
TITLE 1 (bases 1 to 925)
JOURNAL HIRAYAMA,T.
DIRECT SUBMISSION Submitted (01-MAY-1995) to the DDBJ/EMBL/GenBank databases. Takashi
HIRAYAMA,T. Hirayama,T. The Institute of Physical and Chemical Research (RIKEN);
Koyadai 3-1-1, Tsukuba, Ibaraki 305, Japan
(E-mail:hirayama@rctsl.riken.go.jp, Tel:0298-36-4359,
Fax:0298-36-9060)
REFERENCE 2 (bases 1 to 925)
AUTHORS Hirayama,T., Maeda,T., Saito,H. and Shinozaki,K.
TITLE Cloning and characterization of seven cDNAs for
JOURNAL hyperosmolarity-responsive (HOR) genes of Saccharomyces cerevisiae
MEDLINE Mol. Gen. Genet. 249 (2), 127-138 (1995)
FEATURES
source 96086928
location/Qualifiers
    1..925
    /organism="Saccharomyces cerevisiae"
    /strain="RS16"
    /db_xref="taxon:4932"
    17..769
    /gene="Rhr2"
    17..769
    /gene="Rhr2"
    /codon_start=1
    /product="unknown"
    /protein_id="BAA09060.1"
    /db_xref="GI:1236750"
    /translation="MPLTTKPLSLKINAALFVNDGTTIIISOPAIATWRFQKDKPYF
    DAELVHIHSHGWRDYDAIAKFAFDPADEYVNLKEGSIPEKYGHSLEPGAVKLNLA
    LNALPEKEMAVNATSGTDMARKWEDILIKRPEYFTIANVYKQKRPPEYLGKRNGL
    GPDINEDDPSKSKVYVEDAPAGIAACAGCKIVGIATFTFDLFLKEKCDLIYKNH
    ESTIRVGYNMETDEVELIFPDYLYAKDDLKMW"
BASE COUNT      298 a      191 c      176 g      260 t
ORIGIN

alignment_scores:
    Quality: 328.50      Length: 229
    Ratio: 2.313      Gaps: 8
Percent Similarity: 62.009      Percent Identity: 37.555

alignment_block:
US-09-558-284-2 x YSPRHR2C ..

Align seg 1/1 to: YSPRHR2C from: 1 to: 925

10 leupheaspheuaspylthrillevalserthrthvalaiaiaaguly 26
|||||.....:|||||.....:|||||.....:|||||.....:
62 CTATTCGATGTGACGGTACATCATCTCTCAACCGCATTCGCTGC 111
|||||.....:|||||.....:|||||.....:|||||.....:
26 salatrprhlylsleucsytyrclutyrply.....Vala 38
|||||.....:|||||.....:|||||.....:|||||.....:
112 TTTCTGAGCA.....GATTTGCGTAAAGACAAAGCCTTACTTGG 149
|||||.....:|||||.....:|||||.....:|||||.....:
38 spprosergiuleupheliyserrhisserhsglyalaargthrgl 54
|||||.....:|||||.....:|||||.....:|||||.....:
150 ATGCCGAACAGCTTATTCACATCTCTCAGCGTTGGAGAACTTACAGTGC 199
|||||.....:|||||.....:|||||.....:|||||.....:
55 leuargargpherpheprolysleuaspsphraspsnlysglyvalle 71
|||||.....:|||||.....:|||||.....:|||||.....:
200 ATTGCCAAGTGTCTCA.....GACTTTCCTGATGAGAAATACGTTAA 243
|||||.....:|||||.....:|||||.....:|||||.....:
71 ualaleuglulysaspillealahiserryrleuaspthvalserleu 88
|||||.....:|||||.....:|||||.....:|||||.....:
244 CAAGCTAGAAAGTGAATCCAGAAAAGTACGTAACACATCTCATCGAAG 293
|||||.....:|||||.....:|||||.....:|||||.....:

```

```

88 1eprgglYAlaGluAsnLeuLeuSerLeuAspValAspThrGluThr 104
   ::::::::::::::::::::
294 TTCGAGGCTGTCAGTGTGTAATGCTTGAACGCC..... 331
105 GlnLysLysLeuProGluArgLysTrpAlaIleValThrSerGlySerPr 121
   |||||:::|||||:::|||||:::|||||:::
332 .....TTGCCAAGAGAAAAATGGCTGTCGCCACCTCGTACCCG 372
121 oTrLeuAlaPheSerTrpPheGluThrIleLeuLysAsnValGlyLysP 138
   ::|||:::|||||:::|||||:::|||||:::
373 TGACATGGCCAGAAATGCTGCAC...ATTGTAAG...ATCAAGAGAC 416
138 rOlyValPheIleThrGlyPheAspValLysAsnGlyLysProAspPro 154
   ||:::|||||:::|||||:::|||||:::|||||
417 CAGAACTACTATCACCGCCAAATGATGTCAGCAAGTAAAGCTCACCCA 466
155 GluGlyTrpSerArgAlaArgAspLeuLeuArgLysAspLeuGluLeu 171
   |||:::|||||:::|||||:::|||||:::
467 GAAACATACTTAAAGGTGAAACGGTTGGGTTCCCAATTATGACCA 516
171 rGlyLysGlnAspLeuLysTrpValValPheGluAspAlaProValGly 188
   ::|||:::|||||:::|||||:::|||||:::
517 AGACCATTCCAATCTAAAGTGTGCTTTTGAAAGCACGACGCTGGTA 566
188 1eLysAlaGlyLysAlaMetGlyAlaIleThrValGlyIleThrSerSer 204
   |||:::|||||:::|||||:::|||||:::
567 TTGCTGCTGTAGGCTGCTGCTGTAATAATGCTTGTATGCTACCAC 616
205 TyrAspLysSerValLeuPheAspAlaGlyAlaAspTrpValValCys 221
   ::|||:::|||||:::|||||:::|||||
617 TTCGATTGACTTCTTGAAGAAAGGCT.....TGTGA 651
221 pleuThrGlnValSerValValLysAsnAsnGluAsn 233
   |:::|||||:::|||||:::|||||:::
652 CATC.....ATTGTCAGAACCAACGAAATCT 676
seq_name: gb_p13:SCCEM1

seq_documentation_block:
LOCUS SCCEM1 2940 bp DNA PLN 15-OCT-1999
DEFINITION Saccharomyces cerevisiae CEM1 gene.
ACCESSION X73488.1 GI:403310
VERSION beta-ketoacyl synthase: CEM1 gene.
KEYWORDS Saccharomyces cerevisiae.
SOURCE Saccharomyces cerevisiae.
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE
AUTHORS Hartington, A., Herbert, C.J., Tung, B., Getz, G.S. and Slonimski, P.P.
TITLE Identification of a new nuclear gene (CEM1) encoding a protein
homologous to a beta-keto-acyl synthase which is essential for
mitochondrial respiration in Saccharomyces cerevisiae
Mol. Microbiol. 9 (3), 545-555 (1993)
JOURNAL 94018649
MEDLINE 2 (bases 1 to 2940)
REFERENCE Slonimski, P.P.
AUTHORS Direct Submission
TITLE Submitted (21-JUN-1993) P.P. Slonimski, Centre de Genetique
Moleculaire, Centre National de la Recherche Scientif., Avenue de la
Terrasse, Gif sur Yvette, F91198 Cedex, FRANCE
FEATURES
SOURCE 1..2940
Location/Qualifiers
/organism="Saccharomyces cerevisiae"
/strain="CW04"
/db_xref="taxon:4932"
1117..2932
/gene="CEM1"
1117..1121
/gene="CEM1"
/note="putative"
1175..1180
/gene="CEM1"

TATA_signal
1202..1205
/note="putative"
/gene="CEM1"
TATA_signal
1247..1251
/note="putative"
/gene="CEM1"
TATA_signal
1306..1309
/note="putative"
/gene="CEM1"
TATA_signal
1316..1320
/note="putative"
/gene="CEM1"
CDS
1453..2781
/note="putative"
/gene="CEM1"
/codon_start=1
/protein_id="CAB58180.1"
/db_xref="GI:6066682"
/translation="MSRRVITLGLCVTPILGRSLSESWGNLSSKNGLPTYSLPNY
EDYKLRKSIPTITVCKIPENFQENSAIKLFTSODERTSSFKIALRTYEAL
HNAGLINPDITITNLSICLDHFGLISGISTODIYOTSIQFNDNRKRIINPVVK
ILTNAGNVSIRKLRCLSHSVSTACATGNNSTDAFPRLGMDICVAGASTSL
HPLSLAGTRAKSLITTINGISRPFDIQRSGFVUGBCGMIWBSLEHAKRANITSEL
VYGLSDPACHITSPADNGAKRAIEALIMARLEPVDVYVNAHATSTLLGDAEC
LAVASALDGRSKSPILYISSNKAIGHILGAAGVESIFTLISLKDMPHTYLDN
VUTLENNADKLHFIRDKPIVGANPKYALCNFSGGVNTSLFFKMEGS"
2884..2932
/gene="CEM1"
/note="consensus"
BASE COUNT 956 a 594 c 573 g 817 t
ORIGIN
alignment_scores:
Quality: 328.50 Length: 237
Ratio: 2.220 Gaps: 7
Percent Similarity: 62.447 Percent Identity: 35.021
alignment_block:
US-09-558-284-2 x SCCEM1 ..
Align seg 1/1 to: SCCEM1 from: 1 to: 2940
2 AlaGluPheSerAlaAspLeuCysLeuPheAspLeuAspGlyThrIleVal 18
:::|||||:::|||||:::|||||:::
418 TCCTCTTTGAAGTTAAAGCCGCTTGTGACGCTGCACGCTACCATTA 467
18 1serThrThrValAlaAlaGluLysAlaTrpThrLysLeuCysTrpGlu 35
:::|||||:::|||||:::|||||:::
468 CATCTCTCAACACGACCATCTGCTGATCTGAGAG.....GATT 505
35 yrgLy.....ValAspProSerGluLeuPheLysHisSer 46
:::|||||:::|||||:::|||||:::
506 TCGTAAAGCAAAACCTTATTGATGCTGACACGCTTATCAAGTCTCG 555
47 HisGlyValArgThrGlnGluValLeuAlaArgPhePheProLysLeuAs 63
|||||:::|||||:::|||||:::
556 CATGCTTGAGAGAACGTTGATGCATGCTAAGTCTGCTCA.....GA 599
63 pasPThrAspAsnLysGlyValLeuAlaLeuGluLysAspIleAlaHis 80
|:::|||||:::|||||:::|||||:::
600 CTTCGCCAATGAAGAGATGTTAAACAATTAGAAGCTGAATTCGGGTCA 649
80 eryTrLeuAspThrValSerLeuIleProGlyAlaGluAsnLeuLeu 96
:::|||||:::|||||:::|||||:::
650 AGTACGGTGAATAATCCATTGAAGTCCAGGTGAGTTAAGCTGTGCAAC 699
97 serLeuAspValAspThrGluThrGlnLysLysLeuProGluArgLys 113
:::|||||:::|||||:::|||||:::
700 GCTTTGAACGCT.....CTACCAAGAGAAATG 728
113 palAlaIleValThrSerGlySerProTrpLeuAlaPheSerTrpPheGlu 130
|||||:::|||||:::|||||:::

```



```

misc_feature
/clone="HSA67F071_213703_H_sapiens_partial_cDNA_sequence
clone 67F07 strand (-) single read"
complement(8228..10264)
/note="orf, len: 678, CAI:0.11"
/codon_start=1
/protein_id="CA86156.1"
/db_xref="GI:557802"
/translation="MGVHEDDANNTWTEATPGVSSDDCGHRTESIOLQNFNTDM
ESMLDEGRNENSKMLIKRHPLOKFLERVMNVPESDEPSPFPRKMGKIID
FPQPTFKRIPSKILRLILLYCCFMRIFFSLIPLKIPRYHPHMGSKIPILS
LSCNLYLMGECNNEGCKNAKNGPLDNKEVIRIRPALCDRCGWTSAIAGNRVKY
TCYETGGALSEEDPRVYVYPIKSDSPFCASAVAGVYISGCTVSMQAGNSF
PSKGMVNTGFSVAFNSFPQSGFRDIOGGLSCYDPRAAVALNMLFGPIEYLY
DSYGYWINTIVGYWTLVSLDPLTDADHPASVEELFSGTLDLCKGALVYVK
SAVKRTLENGSPIAKVILMYPTFMGLISNNVTFDELPTDRTDLDKROAGALVGS
IAATILTCAVIOAVSLKSGRKPKEYKITYOIGGLIAGSLPGIINLHHYILGSLI
VPGCATRGSSAVILQGLVGLILSGVAREASIVETTTALLRGAGSLKRPILIDN
DDONHSLSMHINATDPVTDQGNIDGFSLLADVEVYGNKMTVSIDVLRNENPLAQ
MMDALDASNGITIDILYKAVASVSPNKRDDYNAGVLQWPNMGKQPEBGVS"
complement(10603..13260)
/gene="f1r3"
complement(10603..13260)
/note="f1r3"
/clone="f1r3, len: 885, CAI: 0.29, conflict with R13_YEAST
P21672_RIBONUCLEOSIDE-DIPHOSPHATE REDUCTASE"
/codon_start=1
/protein_id="CA86157.1"
/db_xref="GI:557803"
/db_xref="SWISS-PROT:P21672"
/translation="MTSKNSSNKSNTPTQWYVIRKDGSRKPEVOFDKITSITLSYG
LDPRIKAVKTOHILISGVSGVTTVELDNLAAFCAMTIVVHEDYATLARIALSNL
HKOTTKQSKVIEDLHMIPNATGKHAMISDELYINVENKDTLNSAIVDRDFOYT
YFGKRTLESLTLKNGVAREPQHLVNRVALGIHSDIESEVLTYNLMSLRPTSHS
PTLENAGTPHPQMSSEFLAMKDSIEGIVDTLKECAMISITAGVGLINNIISTGS
YIAETNGTISNGLIPMIRVFNTARYVDQGGKRGAGALFLEPHADLFDNDVLRKTH
GKEIRARDLEPALMIDLEPMKRVQEDGPPYLFSPSAPGLDVGWGEFEELVRYER
EGSRKTIKAOKLWAILOAOTEGTPEPMYVKADANRKTNOONITIKSSNCCGIVEY
SSPETAVCINASTALPAPVYSPDCKTASNEPRALHEIKATVTHNATVDRNTVY
PEARNSMKRRPILALGVQGLADTYMLRLPESESEDTLNKQIETLYHATVLSCELA
AOKGKSTEGSPASKGILQFDMNNAKPEGMWETLRKDIDYHGGNRSITJVMADPT
ASTQIIGVNECEPVTNMYSRVLSGEEQVWPYLLRDLVDGIDMSKQYLLIQ
NGSIOGLPNVPOLEKELYKTWELISOKTINMADARIDYDOSLSMLPLQAPRMYKI
TSMHYGMKKGLKGTMYLRTQOASAAIOFTIDEVADATTHASVLEDRPYVYK
GTRKEQKASALTESDNEKASPVPSQSSVSNKLEDSVAPVAFETETIKED
SDEKCDIYNEKVIACTAPTPECEGSG"
complement(11392..11427)
/gene="f1r3"
/note="Ribonucleotide reductase large subunit signature"
complement(11410..11472)
/gene="f1r3"
/note="Multicopper oxidases signature 1"
complement(11812..114279)
/note="orf, len: 155, CAI: 0.11"
/codon_start=1
/protein_id="CA86158.1"
/db_xref="GI:557804"

```

```

CDS
/db_xref="SWISS-PROT:P40515"
/translation="MTKYDFWPTLKDAEPLYPQQLLELRQGVSEGGPTATIOSRFN
YAMGLIKSTDVNDERLGVKILITDIKEAESRRECLYLTIGCYKLGESMAKRYVDI
LEFHRNRKVOGALKSWEDKIQKFLTKGVAVAGVLGAVAVASFLRNRR"
14447..15220
/note="orf, len: 257, CAI: 0.13"
/codon_start=1
/protein_id="CA86159.1"
/db_xref="GI:557805"
/translation="MKRSEKSSSALKNGIMERTQPEKVVOMOGTADLSKIGTKK
YWDELALTELEFRNPQDGTGDCWPSDSADKQKIDPLVDNIGAYRISENASVYDLC
GNGHMLPELHETROGKVGDIIDSESKIASNIAEATGVNFIISFOADTIFSGDMKP
GKYDIVLDKGTIDALISGMKINGKLDVYVAGVVERILKIDGIFLTSCNFTODEL

alignment_scores:
      Quality: 328.50      Length: 229
      Ratio: 2.313      Gaps: 8
Percent Similarity: 62.009      Percent Identity: 37.555

alignment_block:
US-09-558-284-2 x SC5610
Align seg 1/1 to: SC5610 from: 1 to: 3730

10 LeupheaspLeuaspGlyThrIleValSerThrValAlaAlaGluLeu 26
|||||.....:|||||.....:
2765 CTATTCGATGCTTGCAGTACATCATCTCTCAACAGCATTCCTGC 27714
26 sAlatpThrLeuLeuGlyTyrGly.....ValA 38
||||
27715 TTTCCTGGACA.....GATTCGGTAAAGCAAGCCTTACTTGC 27752
38 sPProSerGluLeuPheLeuHisSerHisGlyAlaAlaGlyThrGluVal 54
||.....:|||||.....:
27753 ATGCCCAACAGCTTATTCATCTCTCACGCTTGAGAACCTTACATGCC 27802
55 LeuArgArgPhePheProLeuLeuAspThrAspAsnGlyAlaLeu 71
.....:|||||.....:
27803 ATTCGCCAGTGTGCTCCA.....GACTTCTGATGAGAAATACGTAA 27846
71 uAlaLeuGluLeuAspIleAlaHisSerThrLeuAspThrValSerLeu 88
|||||.....:|||||.....:
27847 CAACGTAGAGGTGAAGTAATCCAGAAAGTACGCTGACACCTCATCGAG 27896
88 IeProGlyAlaGluAsnLeuLeuSerLeuAspValAspThrGluThr 104
.....:|||||.....:
27897 TTCACAGTCTCTGTCAAGTGTGTAAATGCTTGAACGCC..... 27934
105 GluLeuLeuLeuProGluArgLysTrpAlaIleValThrSerGlySerPr 121
|||||.....:|||||.....:
27935 .....TTCGCAAGAGAAAATGGGCTGCGCACCTCTGTGTAACCG 27975
27935 .....TTCGCAAGAGAAAATGGGCTGCGCACCTCTGTGTAACCG 27975
121 oTyrlLeuAlaPheSerTrpPheGluThrIleLeuLeuAsnValGlyLysP 138
.....:|||||.....:
27976 TCACATGGCCAGAAATGTTGAC...ATTGTAAG...ATCAGAGAGC 28019
138 rGlyValPheLeuThrGlyPheAspValLysAsnGlyLysTrpAspPro 154
|||||.....:|||||.....:
28020 CAGAACTACTTACACCCCAATGATGTCACAGCAAGTAAGCTCACCA 28069
155 GluGlyTyrSerArgAlaArgAspLeuLeuArgGlnAspLeuLeuThr 171
|||||.....:|||||.....:
28070 GAAACATACCTTAAGGTAGAAACGTTGGTTCCTCCCAATTAAATGAACA 28119
171 rGlyLysGlnAspLeuLysTyrValValPheGlnAspAlaProValGly 188
.....:|||||.....:
28120 AGACCATTCACAAATCTAAGGTGTTGCTTGAAGACCAACACAGCTGTA 28169
188 IeLysAlaGlyLysAlaMetGlyAlaIleThrValGlyIleThrSerSer 204
|||||.....:|||||.....:
28170 TTGCTGCTGTGTAAGGCTGCTGCTGTAAATTCGTTGATTGTAACACT 28219

```


ORIGIN

alignment_scores: Quality: 327.50 Length: 229
 Ratio: 2.323 Gaps: 7
Percent Similarity: 61.572 Percent Identity: 36.245

alignment_block:
US-09-558-284-2 x YSCHORZA ..

Align seg 1/1 to: YSCHORZA from: 1 to: 969

```
10 LeupheaspLeuaspGlyThrIleValSerThrThrValAlaIaIaGlyIuVal 26
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
92 TTGTTCCGACGTCGACGTTACATCTTCATCTCAACCGCCATTCGTCGC 141
26 sAlaTThrThrLeucysTyGluTyGly.....ValA 38
   |||
142 ATCTGGAGG.....GATTTCGTTAGCACAACCTTATTTCG 179
38 sPProSerGluLeuPheLeuHisSerHisGlyAlaArgThrGlnGluVal 54
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
180 ATGCTGACACAGCTATCCAGCTCGCATGCTTGAGAACGTTGATGCC 229
55 LeuArgArgPhePheProLysLeuAspThrAspAsnLysGlyValLe 71
   ::|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
230 ATTGCTAGTTCGCTCA....GACTTTGCCAATGAAGAGTAGTGTAA 273
71 uAlaLeuGluLysAspIleAlaHisSerTyLeuAspThrValSerLeu 88
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
274 CAATTTAGACAGTGAATTCGCGTCACAGTACGCTGAATAATCCATGAA 323
88 lPProGlyAlaGluAsnLeuLeuSerLeuAspValAspThrGluThr 104
   ::|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
324 TCCGAGGTGCACTTAAGCTGTCCAACGCTTGAACGCT..... 361
105 GlnLysLysLeuProGluArgLysThrPalaIleValThrSerGlySerPr 121
362 .....CTACCAAAAGAAAGAAATGGCTGTGGCACTTCGCGTACCCG 402
121 oTyrlLeuAlaPheSerTrpPheGluThrIleLeuLysAsnValGlyLysP 138
   ::|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
403 TCATATGGCACAATAATGGTTCGACATCTG.....GGAATCAGGAGAC 446
138 rOlyValAlaPheIleThrGlyPheAspValLysAsnGlyLysProAspPro 154
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
447 CAAGTACTTCACTACCGCTAATGATGCAACACAGGTAAGCCTCATCCA 496
155 GluGlyTyrserArgAlaArgAspLeuLeuArgGlnAspLeuGlnLeuth 171
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
497 GAACCATATCTGAAGGCGCAATGGCTTAGCATATCCGATCAATGACGA 546
171 rGlyLysGlnAspLeuLysTyrValValPheGluAspAlaProValGlyI 188
   ::|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
547 ACACCTTCCAAATCTAAGTAGTATTGTAAGACGCTCCAGCAGGTA 596
188 lElySalIaGlyLysAlaMetGlyAlaIleThrValGlyIleThrSerSer 204
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
597 TTGCCGCCCGAAAAAGCCGCGTTGTAAGATCAFTGTATGTGCCACTACT 646
205 TyrAspLysSerValLeuPheAspAlaGlyAlaAspTyrValValLysAs 221
   ::|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
647 TTCGACTTGGACTTCTTAAGAAAAAGCC.....TGTGA 681
221 pleuthrGlnValSerValValLysAsnAsnGluAsn 233
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
682 CATC.....ATTGTCAAAAACACAGCAATCC 706
```



```

9  ATGGCAGAAATTTTCAGCTGATCTATGTCCTTTTGACCTAGATGATCAT 58
17  eValSerThrThrValAlaAlaGluLysAlaTrpThrLysLeuCysTyrG 34
59  AGTGAAGTACAAACAGTGGCCGACAGAAACATGAGCAAGATTGGTTAGC 108
34  lUTyG1ValAspProSerGluLeuPheLysHisSerHisGlyAlaArg 50
109  AATAGGGTGTGATCTCTCCGAGTATTTAAGCATTTTCATGTTGACAA 158
51  ThrGlnGluValLeuArgArgPhePheProLysLeuAspPThrAspAs 67
159  ACACAAGAGGTTTGAAGGTTTTCCTAATGTGATGATACAGACAA 208
67  nLysLysValLeuAlaLeuGluLysAspIleAlaHisSerTyrLeuAsp 84
209  TAAAGGTGCTTGGCTAGAAAGAAATATGCCATAGTACTTGGGAA 258
84  hrValSerLeuIleProGlyAlaGluAsnLeuLeuSerLeuAspVal 100
259  CAGTAAGCCTTATCTCGTGCAGAGAACTTACTGTTATCGTTAGATGA 308
101  AspThrGluThrGlnLysLysLeuProGluArgLysTyrPalaIleValTh 117
309  GATACTGAGACTCAAAAAAGTTACTGTAAGGAAATGGGCTATCGTTAC 358
117  rSerLysSerProTyrLeuAlaPheSerTrpPheGluThrIleLeuLys 134
359  CTTCTGGTTCATATTGGCATTTTCATGTTGCGAGACAATATTGAAAA 408
134  snValGlyLysProLysValPheIleThrGlyPheAspValLysAsnGly 150
409  ATGTTGGAAAGCCCAAGTTTCATTACTGGGTTGACCTGGAAGACGGT 458
151  LysProAspProGluGlyTyrSerArgAlaArgAspLeuLeuArgLins 167
459  AAGCTGTGATCCCGAGGGTTATTCAGAGCTCGTGAATTATTCGCTGACA 508
167  PLeuGlnLeuThrGlyLysGlnAspLeuLysTyrValValPheGluAspA 184
509  TTTGCAATTAACTGGTAAACAGAGATCTGAAGTATGTTGCTTCGAAAGATG 558
184  lAprOvalGlyIleLysAlaGlyLysAlaMetGlyValIleThrValGly 200
559  CACCGGTGGGCAATAAGCGCGCAAGCAATGGCGCCATTACTGTGGGT 608
201  lIleThrSerSerTyrAspLysSerValLeuPheAspAlaGlyAlaAsp 217
609  ATTAACATCTCTGATGACAAAGCGGTTTATTTGACGACGAGACAGATTA 658
217  rValValCysAspLeuThrGlnValSerValValLysAsnAsnGluAsnG 234
659  TGTAGTCTGTGATTTGCACACAGGTTTCCGTGTTAAGAAACAAATGAAGA 708
234  lYleValIleGlnValAsnAsnProLeuThrArgAla 246
709  GTATTGTCTCCAGGTAACACACCCCTTGACAAAGGCC 746

seq_name: /SIDS1/gcdata/geneseq/geneseqn/NA200.DAT:AAA13804
seq_documentation_block:
ID AAA13804 standard; DNA; 408 BP.
XX
XX
AC AAA13804;
XX
XX
DE 27-JUL-2000 (first entry)
XX
XX Yeast lacZ nucleotide sequence SEQ ID NO:3.
XX
XX Yeast stress responsive gene; promoter; brewing; beer; wine; sake;
KM bread; oxidative stress; osmotic pressure; stress; glucose starvation;
ds.
XX
XX

```

```

OS Saccharomyces cerevisiae.
XX
XX Key Location/Qualifiers
XX CDS 81..407
XX FT /*tag= a
XX FT /product= "lacZ"
XX FT /note= "no stop codon given"
XX
XX JP2000078977-A.
XX
XX 21-MAR-2000.
XX
XX 04-SEP-1998; 98JP-0251390.
XX
XX 04-SEP-1998; 98JP-0251390.
XX
XX (TAIF ) MARUHA CORP.
XX
XX WPI: 2000-285929/25.
XX
XX P-PSDB; AAY82568.
XX
XX A stress-responsive gene promoter
XX
XX Example 1; Page 9; 12pp; Japanese.
XX
XX
XX The present invention describes a stress responsive gene promoter
XX isolated from Saccharomyces cerevisiae (Yeast). Also described in the
XX present invention are: (1) a promoter containing a DNA hybridising with
XX the above DNA under a stringent condition and having stress-responsive
XX promoter activity; (2) a gene expression cassette containing the above
XX promoter; (3) an expression vector containing the above gene expression
XX cassette; (4) a recombinant vector in which a gene encoding an optional
XX polypeptide is recombined to the above expression vector; (5) a
XX transformant containing the above recombinant vector; and (6) a method
XX for the preparation of the above polypeptide in which the above
XX transformant is cultured and the polypeptide is collected from the
XX resultant culture. Saccharomyces cerevisiae is used for the brewing of
XX beer, wine and sake and production of bread. The gene is responsive to
XX the stresses such as oxidative stress, osmotic pressure stress and
XX glucose starvation stress. The present sequence encodes lacZ, which is
XX used in an example from the present invention.
XX
XX Sequence 408 BP; 147 A; 65 C; 77 G; 119 T; 0 other:
XX
XX
XX alignment_scores:
XX Quality: 478.00 Length: 109
XX Ratio: 4.596 Gaps: 0
XX Percent Similarity: 95.413 Percent Identity: 82.569
XX
XX alignment_block:
XX US-09-558-284-2 x AAA13804 ..
XX
XX Align seg 1/1 to: AAA13804 from: 1 to: 408
XX
XX 1 MetAlaGluPheSerAlaAspLeuCysLeuPheAspLeuAspGlyThrI 17
XX ||| ::::::::::::::::::::::::::::::::::::::::::::::::::::
XX 81 ATGCCACAATTTTCAGTAGATCTTGTGCTTTTGACCTAGATGGGACTAT 130
XX |||
XX 17 eValSerThrThrValAlaAlaGluLysAlaTrpThrLysLeuCysTyrG 34
XX |||
XX 131 TGTCAACACACAACTGCAAGCGGAAAGTGCCTGTAATAAATTGTGCCGTC 180
XX |||
XX 34 lUTyG1ValAspProSerGluLeuPheLysHisSerHisGlyAlaArg 50
XX ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
XX 181 AGCATGGGTTGATCCTGTTGACTTATTCAGGATTCCTCCATGGTGAAGA 230
XX |||
XX 51 ThrGlnGluValLeuArgArgPhePheProLysLeuAspPThrAspAs 67
XX ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
XX 231 TCACAAGAAATGATGAAGAATTTTTCCAAATTTGGACATATCCGATTA 280
XX |||
XX 67 nLysGlyValLeuAlaLeuGluLysAspIleAlaHisSerTyrLeuAsp 84
XX |||
XX

```

281 TAAAGCTGTTGCGTTAGAAAAAGCATATCGCAGATATTATTGACA 330
84 hrValSerLeuIleProGlyAlaGluAsnLeuLeuSerLeuAspVal 100
|||||
331 CAGTAAGCCTTATTCCTGCTGCAGAGAAATTATTATTGTTATCGTTAGATCTA 380
101 AspThrGluThrGlnIleLysLysLeuPro 109
|||||
381 GATACTGACACTCAAAAAAGTTACTT 407
seq_name: /SIDSI/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV42019

seq_documentation_block:
ID AAV42019 standard: DNA; 816 BP.
AC AAV42019;
XX
XX 12-OCT-1998 (first entry)
XX
XX Glycerol-3-phosphatase gpp1 gene.
XX
XX Glycerol-3-phosphatase: gpp1 gene; 1,3-propanediol: ss.
XX
XX Saccharomyces sp.
XX
XX W09821339-A1.
XX
XX 22-MAY-1998.
XX
XX 10-NOV-1997; 97WO-US20292.
XX
XX 13-NOV-1996; 96US-0030601.
XX
XX (DUPO) DU PONT DE NEMOURS & CO E. I.
XX (GEMV) GENENOR INT INC.
XX
XX Dias-Torres M, Gatenby AA, Haynie SL, Hsu AK, Lareau RD;
XX Nagarajan V, Nair RV, Nakamura CE, Payne MS, Picatagalo SK;
XX Trimbur DE, Whited GM;
XX
XX WPI: 1998-297942/26.
XX
XX P-PSDB: AAW30683.
XX
XX Fermentative production of 1,3-propanediol - by single organism
XX containing cassette comprising specific genes, and capable of using
XX inexpensive carbon sources
XX
XX Claim 10; Page 52-53; 95pp; English.
XX
XX This DNA sequence includes the coding region of the gpp1 gene
XX encoding a cytosolic glycerol-3-phosphatase (see AAW30683) of
XX Saccharomyces sp. This enzyme catalyses the conversion of
XX glycerol-3-phosphate to glycerol. A claimed method for production
XX of 1,3-propanediol (1) comprises culturing a microorganism
XX transformed with a cassette containing at least 1 of the genes (see
XX AAV42012-21) for glycerol-3-phosphate dehydrogenase,
XX glycerol-3-phosphatase, glycerol dehydratase and 1,3-propanediol
XX oxidoreductase (see AAW30676-85). Any of these genes not present on
XX the cassette must be present endogenously. The enzyme sequences
XX may include substitutions, deletions and additions provided activity
XX is not altered. A single recombinant organism can now be used for
XX production of (1) from inexpensive C-sources (contrast use of
XX glycerol or dihydroacetone) without causing pollution. (1) is
XX a starting material for polyesters, polyurethanes and cyclic
XX compounds.
XX
XX Sequence 816 BP; 252 A; 178 C; 171 G; 215 T; 0 other;

alignment_scores:
Quality: 328.50 Length: 229
Ratio: 2.313 Gaps: 8
Percent Similarity: 62.009 Percent Identity: 37.555

alignment_block:
US-09-558-284-2 x AAV42019 ..
Align seg 1/1 to: AAV42019 from: 1 to: 816

10 LeuPheAspLeuAspGlyThrIleValSerThrThrValAlaAlaGly 26
|||||
109 CTATTGATGCTTGACGCTACCATCATCATCTCTCAACAGCATTTGCTGC 158
26 SalATrThrLysLeuLysSerGlyThrGly.....ValA 38
|||
159 TTTCTGGAGA.....GATTCGTGAAGACAGCCTTACTTCC 196
38 sPProSerGluLeuPheLysHisSerHisGlyAlaIleArgThrGluVal 54
|||
197 ATGCCGAACAGCTTATTCACATCTCTCAGCGTTGAGAACTTACGATGCC 246
55 LeuArgArgPhePheProLysLeuAspThrAspAsnLysGlyVal 71
::: |||
247 ATTCGAAGTTCCCTCCA.....GACTTTCGTGATGACAGATACGTTAA 290
71 uAlaLeuGluLysAspIleAlaHisSerThrLeuAspThrValSerLeu 88
|||||
291 CAAGCTGAAGGTGAATCCCAAGTAAGGTGAAGTGAACCTCCATCGAG 340
88 lPProGlyAlaGluAsnLeuLeuSerLeuAspValAspThrGluThr 104
::|
341 TTCAGGTGCTGTCAAGTTGTATGCTTTGAAGGCC..... 378
105 GlnLysLysLeuProGluArgLysThrAlaIleValThrSerGlySer 121
|||||
379TTGCCAAGAAAGAAATGGCGTGTCCGACCTGTGTACCG 419
121 cTyrlLeuAlaPheSerThrPheGluThrIleLeuLysAsnValGlyLys 138
::|
420 TGACATGCGCCAGAAATGTTGCAC...ATTTGAG...ATCAAGAGAC 463
138 rOlyValPheIleThrGlyPheAspValLysAsnGlyLysProAspPro 154
|||
464 CAGATATTCTTATCACCAGCCCAATGATGTCAAGCAAGTAAGCTCACCCA 513
155 GlnGlyTyrSerArgAlaArgAspLeuLeuArgGlnAspLeuGlnLeu 171
|||
514 GAACATACCTTAAGGCTAGAAAGCGTTGGTTCCCAATTATTAAGACA 563
171 rGlyLysGlnAspLeuLysTyrValValPheGluAspAlaProValGly 188
::: |||
564 AGACCCATCCAAATCTAAGGTTGTTCTTTGAAGACGACACACCTGGTA 613
188 lEysAlaGlyLysAlaMetGlyAlaIleThrValGlyIleThrSerSer 204
|||
614 TTGCTGCTGGTAGGCTGCTGCGCTGTAATAATCGTTGATATTGCTACACT 663
205 TyrAspLysSerValLeuPheAspAlaGlyAlaAspTyrValCysAs 221
::|
664 TTGATTTTGACCTCTTGAAGAAAGGCT.....TGTGA 698
221 pleuThrGlnValSerValValLysAsnAsnGluAsn 233
|||
699 CATC.....ATTGTCAAGAACCAAGAAATCT 723

seq_name: /SIDSI/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV35740

seq_documentation_block:
ID AAV35740 standard: DNA; 816 BP.
XX
XX AAV35740;
XX
XX

28-SEP-1998 (first entry)
XX
XX Klebsiella pneumoniae glycerol-3-phosphatase GPP1 gene.
XX

KM GPPI gene; glycerol-3-phosphatase; production;
 KM 1,3-propanediol; recombinant; ss.
 XX Klebsiella pneumoniae.
 OS
 PN W09821341-A2.
 XX
 XX 22-MAY-1998.
 PD
 XX 13-NOV-1997; 97WO-US20873.
 XX
 PR 13-NOV-1996; 96US-0030601.
 XX
 PA (GENEV) GENENCOR INT INC.
 XX
 PI Chase MW, Diaz-Torres M, Dunn-Coleman NS, Trimbur D;
 XX WPI: 1998-297944/26.
 DR
 XX
 PT New method for increasing production of 1,3-propanediol - comprises
 PT fermentation of inexpensive carbon sources by microorganism
 PT expressing dehydratase, used, e.g. to prolong half-life of enzyme
 XX
 XX Disclosure: Page 62; 133pp; English.
 PS
 XX The sequence is that of the GPPI gene which encodes cytosolic
 CC glycerol-3-phosphatase. It was used as part of a method of
 CC fermentative production of 1,3-propanediol (1,3-pd), using an organism
 CC comprising at least 1 gene encoding a dehydratase, is improved by
 CC inserting into the host a gene encoding protein X and culturing
 CC the transformant in presence of a carbon source (e.g. mono-, oligo-
 CC or poly-saccharide or 1C substrate) convertible to 1,3-pd.
 CC 1,3-pd is a starting material for polyesters, polyurethanes and
 CC cyclic compounds. 1,3-pd can now be produced by a single
 CC recombinant organism from inexpensive carbon sources such
 CC as glucose (rather than costly glycerol or dihydroxyacetone),
 CC rapidly and without causing pollution.
 CC
 XX Sequence 816 BP; 252 A; 178 C; 171 G; 215 T; 0 other:
 XX
 XX
 alignment_scores:
 Quality: 328.50 Length: 229
 Ratio: 2.313 Gaps: 8
 Percent Similarity: 62.009 Percent Identity: 37.555
 alignment_block:
 US-09-558-284-2 x AAV35740 ..
 Align seg 1/1 to: AAV35740 from: 1 to: 816
 10 LeupheaspLeuaspDgIYThrlleValSerThrValAlaAlaGluIy 26
 |||||.....
 109 CTAITCGATGTTGACGTCATCATCTCTCAACACGCACTTGCTGC 158
 26 salatrprhrlLysLeuGysTyrGluIy.....ValA 38
 |||||.....
 159 TTTCTGAGAA.....GATTGCGTAAGACAGACGCTTACTTGC 196
 38 spprSerGluLeuPheLysHisSerHisGlyAlaArgThrGlnGluVal 54
 |||||.....
 197 ATGCCGAACGTTATTCATCTCTCAGCGTTGGAGAACTTACGATGCC 246
 55 LeuAcrGArPhePheProLysLeuAspAspThrAspAsnLysGlyAla 71
 |||||.....
 247 ATTGCCAGATGCGTCCA.....GACTTGTCTGATGAAGAAATACGTTAA 290
 71 uAlaLeuGluLysAspIleAlaHisSerTyrLeuAspThrValSerLeu 88
 |||||.....
 291 CAAGCTAGAGAGTGAATCCAGAAAGTACGATGAGACCTCAGTGAAG 340
 88 leProGlyAlaGluAsnLeuLeuSerLeuAspValAspThrGlnThr 104
 |||||.....

341 TTCCAGGTGCTGTCAGATTTGTATATGCTTTGAACGCC..... 378
 105 GlnLysLysLeuProGluArgLysTrpAlaIleValThrSerGlySerPr 121
 |||||.....
 379TTGCCAAAGGAAAAATGGGCTGTGCCACCTCTGTACCCG 419
 121 cTyrLeuAlaPheSerTrpPheGluThrIleLeuLysAsnValGlyLysP 138
 |||||.....
 420 TGACATG6CCAGAAAGTTGAC...ATTTTGAAG...ATCAAGAGAC 463
 138 rOLysValPheIleThrGlyPheAspValLysAsnGlyLysProAspPro 154
 |||||.....
 464 CAGAAATACCTCATCACCGCAATGATGTCAAGCAGTAAGCTCACCCA 513
 155 GluGlyTyrSerArgAlaArgAspLeuLeuArgGlnAspLeuGlnLeuTh 171
 |||||.....
 514 GAACCATACTTAAGGAGTAAAGGTTGGCTTCCCATTAATATGAGACA 563
 171 rGlyLysGlnAspLeuLysTyrValValPheGluAspAlaProValGlyI 188
 |||||.....
 564 AGACCCATCAATCTTAAGGTTGTGCTTTGAGACGACACAGTGTGA 613
 188 leLysAlaGlyLysAlaMetGlyAlaIleThrValGlyIleThrSerSer 204
 |||||.....
 614 TTGCTGCTGTTAAGGCTGCTGCTGTAATAATGCTGTGTAATGCTACACT 663
 205 TyrAspLysSerValLeuPheAspAlaGlyAlaAspTyrValValLysAs 221
 |||||.....
 664 TTCGATTTGGACTCTTGAAGGAAAGGTT.....TGTGA 698
 221 pleuThrGlnValSerValValLysAsnAsnGluAsn 233
 |||||.....
 699 CATC.....ATTGTCAGAGAACCAAGATCT 723
 seq_name: /SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV30825
 seq_documentation_block:
 ID AAV30825 standard; DNA; 816 BP.
 XX
 AC AAV30825;
 XX
 DT 14-SEP-1998 (first entry)
 XX
 DE Saccharomyces cytosolic glycerol-3-phosphatase GPPI gene.
 XX
 KW Glycerol-3-phosphatase; G3P; GPPI; RHR2; YIL053W; ss.
 XX
 OS Saccharomyces sp.
 XX
 PN W09821340-A1.
 XX
 PD 22-MAY-1998.
 XX
 PF 10-NOV-1997; 97WO-US20293.
 XX
 PR 13-NOV-1996; 96US-0030602.
 XX
 PA (DUPLO) DU PONT DE NEMOURS & CO E I.
 PA (GENEV) GENENCOR INT INC.
 XX
 PI Butlhis BA, Gatenby AA, Haynie SL, Hsu AK, Lareau RD;
 XX WPI: 1998-297943/26.
 DR P-PSDB: AAW57327.
 XX
 PT Fermentative production of glycerol using recombinant host -
 PT containing genes for glycerol-3-phosphate dehydrogenase and/or
 PT glycerol-3-phosphatase
 XX
 PS Disclosure: Page 28; 57pp; English.
 XX
 CC This Saccharomyces sp. gene, designated GPPI, RHR2 or YIL053W,
 CC codes for a cytosolic glycerol-3-phosphatase (G3P, see AAW57327) that

CC catalyses the conversion of glycerol-3-phosphate to glycerol.
CC The invention provides recombinant organisms that express G3P and/or
CC glycerol-3-phosphate dehydrogenase (G3PDH) (see also AAM57324-32)
CC useful for the production of glycerol from a variety of C-sources.
CC A host cell is preferably transformed with a cassette containing
CC either a G3PDH gene and/or a G3P gene and then cultured in the
CC presence of a mono-, oligo-, polysaccharide or IC-substrate. The
CC glycerol obtained is used in cosmetics, liquid soaps, pharmaceuticals,
CC lubricants and antifreezes; its esters are used in the oil and fat
CC industries. The method produces glycerol rapidly and inexpensively
CC without generation of polluting by-products.

XX Sequence 816 BP; 252 A; 178 C; 171 G; 215 T; 0 other:

alignment_scores:

Quality:	328.50	Length:	229
Ratio:	2.313	Gaps:	8
Percent Similarity:	62.009	Percent Identity:	37.555

alignment_block:

US-09-558-284-2 x AAV30825

Align seg 1/1 to: AAV30825 from: 1 to: 816

```
10 LeupheaspleuaspglyThrIleValSerThrValAlaIaIaGlyIuLy 26
|||||
109 CTATTCAGTGTGACGAGTACATCATCTCTCACACGACCATTCGCTGC 158
26 SAlaTrpThrIuLySerCysTyrgIuTyrgIy.....ValA 38
|||||
159 TTTCGTGGAGA.....GATTCGTGAAGACAAAGCCTTACTTCG 196
38 spproSerGluLeuPheIuHisSerHisGlyAlaargThrGlnIuLy 54
|||||
197 ATGCCGAACAGCTTATTCACATCTCTCACGCTTGAGAACTTACATGCC 246
55 LeuArgArpPheProLysLeuAspThrAspAsnLysGlyValle 71
|||||
247 ATTCGCAAGTTCGCTCCA.....GACTTGTCTGATGAAGATACGTTAA 250
71 uAlaLeuGluLysAspIleAlaHisSerTygLeuAspThrValSerLeu 88
|||||
291 CAACTGGAAGAGTAAATCCCAAGAAAGTACGAGGACCTCATCGAAG 340
88 lEProGlyAlaGluAsnLeuLeuSerLeuAspValAspThrGlnIuThr 104
|||||
341 TTCAGGTGCTGTCAAGTTGTATGCTTTGAACGCC..... 378
105 GluIuLysLeuProGluArgLysTrpAlaIleValThrSerGlySerPr 121
|||||
379 .....TTGCCAAGAGAAATAATGGCGCTGCCACCTGTGGTACCG 419
121 cTyLeuAlaPheSerTrpPheGluThrIleLeuLysAsnValGlyLysP 138
|||||
420 TGACATCGCCCAAGAAATGCTTCGAC...ATTTGAAG...ATCAAGAGC 463
138 rOlusValPheIleThrGlyPheAspValLysAsnGlyLysProAspPro 154
|||||
464 CAGAAATCTTATCACCAGCATATGTCAGCAAGCAAGCAACCTCACCCA 513
155 GluIuLysSerArgAlaArgAspLeuLeuArgGlnAspLeuGlnIuLeuTh 171
|||||
514 GAACCATACTTAAGGGTACAAACGGTTGGTTCCCAATTAATGAACA 563
171 rGlyLysGlnAspLeuLysTygValValPheGluAspAlaProValGlyT 188
|||||
564 AGACCCATCCAAATCTAAAGCTTGTCTTTGAAGACCCACACCTGGTA 613
188 lElysAlaGlyLysAlaMetGlyAlaIleThrValGlyIleThrSerSer 204
|||||
614 TTGCTGCTGGTAAGCTGCTGCTGTAATAATCGTTGCTATTGCTACACT 663
```

205 TyAspLysSerValLeuPheAspAlaGlyAlaAspTygValValCysAs 221
:::|||||
664 TTGCATTTGGACTTCTTGGAGCAAAAGCTTGTCGA 698
221 PleuThrGlnValSerValValLysAsnGlnLysN 233
:::|||||
699 CATC.....ATTGTCMAAGACCACCAACT 723
seq_name: /SIDSL/gcdata/geneseq/geneseq/NM1999.DAT:AAx80614

seq_documentation_block:

ID AAX80614 standard; DNA; 816 BP.

XX AAX80614;

XX 29-SEP-1999 (first entry)

XX GPP1 gene.

XX GPP1: glycerol-3-phosphatase; glycerol; recombinant organism;

KW transformation; glycerol biosynthetic pathway; expression cassette;

KW 1-3-propanediol; pharmaceutical compound; antifreeze solution;

XX lubricant; polyurethane; cyclic compound; fat and oil industry;

XX polyester fiber; ss.

XX Saccharomyces sp.

XX Key

XX 10-JUN-1999.

XX 02-DEC-1998; 98WO-US25551.

XX 02-DEC-1997; 97US-0982783.

XX (DUPD) DU PONT DE NEMOURS & CO E. I.

XX (GEMV) GENENCOR INT INC.

XX NaIr RV, Payne MS, Trimbur DE, Valle F;

XX WPI: 1999-385384/32.

XX P-PSDB: AAY26169.

XX Recombinant organisms containing G3PDH and or G3P phosphatase

XX Disclosure; Page 55; 84pp; English.

XX The present sequence is a GPP1 gene encoding cytosolic glycerol-3-

XX phosphatase which catalyses the conversion of glycerol-3-phosphate and

XX water to glycerol and inorganic phosphate. GPP1 is alternatively

XX referred to as RHR2 or YII053W. This is used to produce glycerol from a

XX recombinant organism by transforming a suitable host cell with an

XX expression cassette comprising either one or both of the genes encoding

XX G3PDH and G3P, where the host cell has disruptions in either glycerol

XX kinase or glycerol dehydrogenase endogenous genes to prevent their

XX active expression. The transformed host cell is cultured with a carbon

XX source and glycerol is recovered. Compounds derived from the glycerol

XX biosynthetic pathway like 1,3-propanediol can also be produced. The

XX method provides a rapid, inexpensive and environment-friendly source of

XX glycerol. Glycerol is used in cosmetics, food, pharmaceuticals,

XX lubricants, anti-freeze solutions, fat and oil industry etc.. 1,3

XX -propanediol is used for the production of polyester fibers and the

XX manufacture of polyurethanes and cyclic compounds.

XX Sequence 816 BP; 252 A; 178 C; 171 G; 215 T; 0 other:

alignment_scores:

Quality: 328.50 Length: 229
Ratio: 2.313 Gaps: 8
Percent Similarity: 62.009 Percent Identity: 37.555

alignment_block:

US-09-558-284-2 x AAX80614 ..

Align seg 1/1 to: AAX80614 from: 1 to: 816

```

10 LeupheaspLeuaspGlyThrIleValSerThrThrValAlaIaIaGlu 26
   |||.....|.....|.....|.....|.....|.....|.....|
109 CTATTCGATGTGACGATACATCATCATCTCTCAACACGACCATGCTGC 158
26 sAlatrrpThrLysLeucysTyrgluTyrgly.....ValA 38
   |||.....|.....|.....|.....|.....|.....|.....|
159 TTTCTGGAGA.....GATTTCGTAAGACAAAGCCTTACTTCG 196
38 spproSerGluLeupheLysHisSerHisGlyAlaIaIaIaIaIaIa 54
   |||.....|.....|.....|.....|.....|.....|.....|
197 ATGCCGAACACGTTATTCACATCTCTCACGCGTTGGAGAACTTACGATGCC 246
55 LeuArGArpHePheProLysLeuaspAspThrAspAsnLysGlyVal 71
   ::::|.....|.....|.....|.....|.....|.....|.....|
247 ATTGCCAAGTTCGCTCA.....GACTTTCGTGATGAGAAATACGTAA 290
71 uAlaIeugIuLysAspIleAlaHisSerTyrrLeuaspThrValSerLeu 88
   |||.....|.....|.....|.....|.....|.....|.....|
291 CAAGCTAGAGGTAAGAAATCCAGAAAGTACGATGAAACATCCATCGANG 340
88 leProGlyValaIaIuaspLeuLeuLeuSerLeuaspValaIaIaIa 104
   ::::|.....|.....|.....|.....|.....|.....|.....|
341 TTCACAGTCTCTCAAGTGTGTAATGCTTGACGCC..... 378
105 GluLysLysLeuProGluArGlyStrPaIaIaIaIaIaIaIaIaIa 121
   |||.....|.....|.....|.....|.....|.....|.....|
379 .....TTGCCAAAGCAAAATGGCTGTCGCCACCTCGTGTACCG 419
121 ofTyrrLeuAlaPheSerTrpPheGluThrIleLeuLysAsnValGlyLysP 138
   |||.....|.....|.....|.....|.....|.....|.....|
420 TCACATAGGCCAAGAAATGCTGCAC...ATTTTGAG...ATCAAGAGAC 463
138 folysValaPheIleThrGlyPheaspValaLysAsnGlyLysProAsp 154
   |||.....|.....|.....|.....|.....|.....|.....|
464 CAGAAATACTTATCAACCCCAATGATGTCACAGCAAGTAAAGCTCACCA 513
155 GluGlyTyrrSerArGAlaArGaspLeuLeuArGlnAspLeuGlnLeu 171
   |||.....|.....|.....|.....|.....|.....|.....|
514 GAACCATACTTAAGGGTAGAAACGGTTGGCTTCCCAATTAATAGACA 563
171 rGlyLysGlnAspLeuLysTyrrValaIaPheGluaspAlaProValGly 188
   ::::|.....|.....|.....|.....|.....|.....|.....|
564 AGACCATCAAAATCTAAGTGTGTTCTTTTGAAGACGACACCGCTGGTA 613
188 leLysAlaGlyLysAlaMetGlyAlaIleThrValGlyIleThrSerSer 204
   |||.....|.....|.....|.....|.....|.....|.....|
614 TTGCTGCTGTGAAGCGCTGCTGCTGAATAATCGTGTGTAATTCCTACAC 663
205 TyrrAspLysSerValLeupheaspAlaGlyAlaAspTyrrValaLysAs 221
   |||.....|.....|.....|.....|.....|.....|.....|
664 TTCGATTGTGACTTCTTGAAGGAAAGGGT.....TGTGA 698
221 pleuThrGlnValSerValaIaLysAsnGlnuAsn 233
   |||.....|.....|.....|.....|.....|.....|.....|
699 CATC.....ATTGTCAGAAACACGAAATCT 723

```

seq_name: /SIDS1/gcgdata/geneseq/geneseq/NAL98.DAT:AAV42020

seq_documentation_block:

ID AAV42020 standard; DNA: 753 BP.

XX AAV42020;

XX

DT 12-OCT-1998 (first entry)

XX Glycerol-3-phosphatase gpp2 gene.

XX Glycerol-3-phosphatase; gpp2 gene; 1,3-propanediol; ss.

XX Saccharomyces cerevisiae.

XX WO9821339-A1.

XX 22-MAY-1998.

XX 10-NOV-1997; 97WO-US20292.

XX 13-NOV-1996; 96US-0030601.

XX (DUPO) DU PONT DE NEMOURS & CO E I.

XX (GEMV) GENENCOR INT INC.

XX Dias-Torres M, Gatenby AA, Haynie SL, Hsu AK, Lareau RD;

XX NagaraJan V, Nair RV, Nakamura CE, Payne MS, Picataccio SK;

XX Trimbur DE, Whited GM;

XX WPI; 1998-297942/26.

XX P-PSDB; AAW30684.

XX Fermentative production of 1,3-propanediol - by single organism

XX containing cassette comprising specific genes, and capable of using

XX inexpensive carbon sources

XX Claim 10; Page 53; 95pp; English.

XX This DNA sequence includes the coding region of the gpp2 gene

XX encoding a cytosolic glycerol-3-phosphatase (see AAW30684) of

XX Saccharomyces cerevisiae. This enzyme catalyses the conversion of

XX glycerol-3-phosphate to glycerol. A claimed method for production

XX of 1,3-propanediol (1) comprises culturing a microorganism

XX transformed with a cassette containing at least 1 of the genes (see

XX AAW42012-21) for glycerol-3-phosphate dehydrogenase,

XX glycerol-3-phosphatase, glycerol dehydratase and 1,3-propanediol

XX oxidoreductase (see AAW30676-85). Any of these genes not present on

XX the cassette must be present endogenously. The enzyme sequences

XX may include substitutions, deletions and additions provided activity

XX is not altered. A single recombinant organism can now be used for

XX production of (1) from inexpensive C-sources (contrast use of

XX glycerol or dihydroxyacetone) without causing pollution. (1) is

XX a starting material for polyesters, polyurethanes and cyclic

XX compounds.

XX Sequence 753 BP; 225 A; 162 C; 173 G; 193 T; 0 other;

alignment_scores:

Quality: 327.50 Length: 229
Ratio: 2.323 Gaps: 7
Percent Similarity: 61.572 Percent Identity: 36.245

alignment_block:

US-09-558-284-2 x AAV42020 ..

Align seg 1/1 to: AAV42020 from: 1 to: 753

```

10 LeupheaspLeuaspGlyThrIleValSerThrThrValAlaIaIaGlu 26
   |||.....|.....|.....|.....|.....|.....|.....|
46 TTGTCGACGTCGACGATACATCATCATCTCTCAACACGACCATGCTGC 95
26 sAlatrrpThrLysLeucysTyrgluTyrgly.....ValA 38
   |||.....|.....|.....|.....|.....|.....|.....|
96 ATTCTGGAGG.....GATTTCGTAAGACAAACCTTATTTCG 133
38 spproSerGluLeupheLysHisSerHisGlyAlaIaIaIaIaIaIa 54
   |||.....|.....|.....|.....|.....|.....|.....|
134 ATGCTGAACACGTTATCCAGTCTCGCATGTGCGATGAGAACGTTGATGCC 183

```

```

55 LeuAArgPhePheProLysLeuAspSphThraspaNlysglyValle 71
   ::::: ||| ||| ::::: |||
184 ATTGCTAAGTTCGCTCCA.....GACTTTCCAATGAAGATGTTAA 227
71 uAlaLeuGluLysAspIleAlaHisSerTyrLeuAspThrValSerLeu 8
   ||||| ::::: ||| ::::: |||
228 CAAATTGAGAGCTGAATTCGGTCAAGTACGGTGAATAATCCATTGAAG 277
88 LeProGlyAlaGluAsnLeuLeuSerLeuAspValAspThrGluThr 104
   ::::: ||| ||| ::::: |||
278 TCCAGGTGCGAGTTAAGCTGTGCAACGCTTGGAACGCT..... 315
105 GlnLysLysLeuProGluArgLysThrAlaIleValThrSerGlySerPr 121
   ||||| ::::: ||| ||| ::::: |||
316 .....CTACCAAAAGAGAAATGGCGTGTGGCAACTTCGGTACCG 356
121 cTyrLeuAlaPheSerThrPheGluThrIleLeuLysAsnValGlyLysP 138
   ::::: ||| ||| ::::: |||
357 TGATATGGCACAAATAATGTTGAGCATCTG.....GGAATCAGAGAC 400
138 rGlyValPheIleThrGlyPheAspValLysAsnGlyLysProAspPro 154
   ||||| ::::: ||| ||| ::::: |||
401 CAAAGTACTTCATTACCGCTAATATGTCAAACAGCGGTAACCTTCATCCA 450
155 GlnGlyTyrSerArgAlaArgAspLeuLeuArgGlnAspLeuGlnLeuTh 171
   ||| ||| ::::: ||| ||| ::::: |||
451 GAACCATATCTGAAGGCGAGCATGCTTACGATATCCGATCAATGACCA 500
171 rGlyLysGlnAspLeuLysTyrValValPheGluAspAlaProValGlyI 188
   ::::: ||| ||| ||||| ::::: |||
501 AGACCTCTCAATCTAAGCTAGTATTGTAAGACGCTCCACAGCTA 550
188 LeuAlaGlyLysAlaMerGlyValIleThrValGlyIleThrSerSer 204
   ||| ||||| ||||| |||||
551 TTGCGCGCGGAAAGCGCGGTGTGAAGATCATGTGATTCGCCACTACT 600
205 TyrAspLysSerValLeuPheAspAlaGlyAlaAspTyrValValCysAs 221
   ::::: ||| ||| ::::: |||
601 TTGCACTTGGACTTCTTAAGCAAAAGGC.....TGGA 635
221 PleuThrGlnValSerValValLysAsnAsnGluAsn 233
   ::::: ||| ||||| ::::: |||
636 CATC.....ATTGTCAAAACCAAGAAATCC 660
seq_name: /SIDSI/gcdata/geneseq/geneseqn/NA1998.DAT.AAV35741
seq_documentation_block:
ID AAV35741 standard; DNA; 753 BP.
XX
XX AAV35741:
XX
XX 28-SEP-1998 (first entry)
XX
XX Klebsiella pneumoniae glycerol-3-phosphatase GPP2 gene.
DE
XX
XX GPP2 gene; glycerol-3-phosphatase; production;
KM 1,3-propanediol; recombinant; ss.
XX
XX Klebsiella pneumoniae.
OS
XX
XX Key Location/Qualifiers
FH 1..753
FT CDS /*tag= a
FT /product= glycerol-3-phosphatase
XX
XX W09821341-A2.
XX
XX 22-MAY-1998.
XX
XX 13-NOV-1997; 97WO-US20873.
XX
XX 13-NOV-1996; 96US-0030601.

```

```

XX
XX (GENEV ) GENENCOR INT INC.
XX
XX PI Chase MW, Diaz-torres M, Dunn-coleman NS, Trimbur D;
XX
XX DR WPI, 1998-297944/26.
XX
XX DR P-PSDB; AAW60261.
XX
XX PT New method for increasing production of 1,3-propanediol - comprises
XX PT fermentation of inexpensive carbon sources by microorganism
XX PT expressing dehydratase, used, e.g. to prolong half-life of enzyme
XX
XX PS Disclosure; Page 62-63; 133pp: English.
XX
XX The sequence is that of the GPP2 gene which encodes cytosolic
XX glycerol-3-phosphatase. It was used as part of a method of
XX fermentative production of 1,3-propanediol (1,3-pd), using an organism
XX comprising at least 1 gene encoding a dehydratase, is improved by
XX inserting into the host a gene encoding protein X and culturing
XX the transformant in presence of a carbon source (e.g. mono-, oligo-
XX or poly-saccharide or 1C substrate) convertible to 1,3-pd.
XX 1,3-pd is a starting material for polyesters, polyurethanes and
XX cyclic compounds. 1,3-pd can now be produced by a single
XX recombinant organism from inexpensive carbon sources such
XX as glucose (rather than costly glycerol or dihydroxyacetone),
XX rapidly and without causing pollution.
XX
XX SO Sequence 753 BP; 225 A; 162 C; 173 G; 193 T; 0 other.

```

```

alignment_scores:
Quality: 327.50 Length: 229
Ratio: 2.323 Gaps: 7
Percent Similarity: 61.572 Percent Identity: 36.245

```

```

alignment_block:
US-09-558-284-2 x AAV35741 ..

```

```

Align seg 1/1 to: AAV35741 from: 1 to: 753

```

```

10 LeuPheAspLeuAspGlyThrIleValSerThrThrValAlaIleGlyL 26
   ||||| ::::: ||| ||| ::::: |||
46 TTGTCGACGTGACGATACCATTCATCTCTCAACAGCATTTGCTGC 95
26 sAlaThrPheLysLeuCysTyrGluTyrGly.....ValA 38
   |||
96 ATTCTGAGAG.....GATTTCGTTAAGGACAAACCTTATTTCG 133
38 sProSerGluLeuPheLysHisSerHisGlyAlaArgThrGlnGluVal 54
   ||| ::::: ||| ||||| |||||
134 ATGCTGAACAGCTTATCCAGCTTCGCGTGTGGAGACGTTTGATGCC 183
55 LeuAArgPhePheProLysLeuAspSphThraspaNlysglyValle 71
   ::::: ||| ||| ::::: |||
184 ATTGCTAAGTTCGCTCCA.....GACTTTCCAATGAAGATGTTAA 227
71 uAlaLeuGluLysAspIleAlaHisSerTyrLeuAspThrValSerLeu 8
   ||||| ::::: ||| ::::: |||
228 CAAATTGAGAGCTGAATTCGGTCAAGTACGGTGAATAATCCATTGAAG 277
88 LeProGlyAlaGluAsnLeuLeuSerLeuAspValAspThrGluThr 104
   ::::: ||| ||| ::::: |||
278 TCCAGGTGCGAGTTAAGCTGTGCAACGCTTGGAACGCT..... 315
105 GlnLysLysLeuProGluArgLysThrAlaIleValThrSerGlySerPr 121
   ||||| ::::: ||| ||| ::::: |||
316 .....CTACCAAAAGAGAAATGGCGTGTGGCAACTTCGGTACCG 356
121 cTyrLeuAlaPheSerThrPheGluThrIleLeuLysAsnValGlyLysP 138
   ::::: ||| ||| ::::: |||
357 TGATATGGCACAAATAATGTTGAGCATCTG.....GGAATCAGAGAC 400
138 rGlyValPheIleThrGlyPheAspValLysAsnGlyLysProAspPro 154

```


seq_documentation_block:
ID AAX80615 standard; DNA: 753 BP.
XX
AC AAX80615.
XX
DT 29-SEP-1999 (first entry)
XX
DE GPP2 gene.
XX
KM GPP2: glycerol-3-phosphatase; glycerol; recombinant organism;
KM transformation; glycerol biosynthetic pathway; expression cassette;
KM 1-3 propanediol; pharmaceutical compound; antifreeze solution;
KM lubricant; polyurethane; cyclic compound; fat and oil industry;
KM polyester fiber; EC 3.1.3.-; ss.
XX
OS Saccharomycetes cerevisiae.
XX
FH key Location/Qualifiers
FT 1..753
FT /*tag= a
FT /product= "glycerol-3-phosphatase"
FT /note= "cytosolic"
XX
PN MO9928480-A1.
XX
PD 10-JUN-1999.
XX
PF 02-DEC-1998: 98MO-US25551.
XX
PR 02-DEC-1997: 97US-0982783.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E. I.
PA (GEMV) GENENCOR INT INC.
XX
PI Nair RV, Payne MS, Trimbur DE, Valle F:
XX
DR WP1: 1999-385384/32.
DR P-SDB: AAY26170.
XX
PT Recombinant organisms containing G3PDH and or G3P phosphatase
XX
PS
PS Disclosure: Pages 55-56; 84pp; English.
XX
CC The present sequence is a Gpp2 gene encoding cytosolic glycerol-3-
CC phosphatase which catalyses the conversion of glycerol-3-phosphate and
CC water to glycerol and inorganic phosphate. Gpp2 is alternatively
CC referred to as HOR2 or YER062C. This is used to produce glycerol from a
CC recombinant organism by transforming a suitable host cell with an
CC expression cassette comprising either one or both of the genes encoding
CC G3PDH and G3P, where the host cell has disruptions in either glycerol
CC kinase or glycerol dehydrogenase endogenous genes to prevent their
CC active expression. The transformed host cell is cultured with a carbon
CC source and glycerol is recovered. Compounds derived from the glycerol
CC biosynthetic pathway like 1,3-propanediol can also be produced. The
CC method provides a rapid, inexpensive and environment-friendly source of
CC glycerol. Glycerol is used in cosmetics, food, pharmaceuticals,
CC lubricants, anti-freeze solutions, fat and oil industry etc.. 1,3
CC -propanediol is used for the production of polyester fibers and the
CC manufacture of polyurethanes and cyclic compounds.
XX
SQ Sequence 753 BP: 225 A; 162 C; 173 G; 193 T; 0 other:

alignment_scores:
Quality: 327.50 Length: 229
Ratio: 2.323 Gaps: 7
Percent Similarity: 61.572 Percent Identity: 36.245

alignment_block:
US-09-558-284-2 x AAX80615 ..

Align seg 1/1 to: AAX80615 from: 1 to: 753

10 LeuPheAspLeuAspGlyThrIleValSerThrThrValAlaIleGly 26
|||||
46 TTGCTGACGTCGACGTCATCTATCATCTCAACGACCATTCGTCG 95
26 sAlatrPhrIleCysTyrGluTyrGly.....ValA 38
|||
96 ATTCTGAGG.....GATTCGGTAAGGACAAACCTTATTTCG 133
38 sPProSerGluLeuPheLysHisSerHisGlyAlaArgThrGlnGluVal 54
||
134 ATGCTGAAACGCTTATCCAGTCGTCATGTTGGAGAACGTTTATGTC 183
55 LeuArgArgPhePheProLysLeuAspThrAspAsnLysGlyValIle 71
|||
184 ATTGCTAAGTTCGTCACA.....GACTTGGCAATGAAGATATGTTAA 227
71 uAlaLeuGluLysAspIleAlaHisSerTyrLeuAspThrValSerLeuI 88
|||||
228 CAATTTAGAAAGCTGAATTCGCGTCAGTACGTAAGTAATCATTTAG 277
88 IeProGlyAlaGluAsnLeuLeuSerLeuAspValAspThrGluThr 104
|||
278 TCACAGGTGCACTTAAGCTGTGCAACGCTTTGACGCT..... 315
105 GlnLysLysLeuProGluArgLysTrpAlaIleValThrSerGlySerPr 121
|||||
316CTACCAAAAGAGAAATGGCGTGGCACTCCGATCCG 356
121 oTyrLeuAlaPheSerTrpPheGluThrIleLeuLysAsnValGlyLysP 138
|||
357 TGATATGGCACAAATAATGTTGACATCTG.....GGAATCAGAGAC 400
138 rGlyValPheIleThrGlyPheAspValLysAsnGlyLysProAspPro 154
|||||
401 CAAGTACTTCAATTACCGTAAATGATGTAACAGCGTAAGCTCATCA 450
155 GluGlyTyrSerArgAlaArgAspLeuLeuArgGlnAspLeuGluLeuThr 171
|||
451 GAACCATATCTGAAGGACGAGAAATGGCTTAGATATCCGATCAATCAGCA 500
171 rGlyLysGlnAspLeuLysTyrValValPheGluAspAlaProValGlyI 188
|||
501 AGACCTTCCTCAATCTAAGGTAGTATTGAAAGACGCTCCAGCAGTA 550
188 IeLysAlaGlyLysAlaMetGlyAlaIleThrValGlyIleThrSerSer 204
||
551 TTGCCCGCCGAAAGCCGCGGTTGTAAGATCATTTGATTCACACTACT 600
205 TyrAspLysSerValLeuPheAspThrAlaGlyAlaAspTyrValValCysAs 221
601 TTCGACTTGACTTCCTTAAGCAAAAGGC.....TGTA 635
221 PleuThrGlnValSerValValLysAsnAsnGlnAsn 233
|||
636 CATC.....ATTGTCAAAAACCCAGCAATCC 660

seq_name: /SIS1/gcdata/geneseq/geneseqn/NM1999.DAT: AAX12984

seq_documentation_block:
ID AAX12984 standard; DNA: 17087 BP.
XX
AC AAX12984:
XX
DT 19-MAR-1999 (first entry)
XX
DE Enterococcus faecalis genome contig SEQ ID NO:47.
XX
KM Enterococcus faecalis; contig; detection; Enterococcal infection;
KM vaccine; attenuation; computer readable medium; ds.
XX
OS Enterococcus faecalis.

```

PN  W09850555-A2.
XX
PD  12-NOV-1998.
XX
PF  04-MAY-1998; 98WO-US08985.
XX
PR  14-NOV-1997; 97US-0066009.
PR  06-MAY-1997; 97US-0044031.
PR  16-MAY-1997; 97US-0046655.
XX
PA  (HUMA-) HUMAN GENOME SCI INC.
PI  Barash SC, Dillon PJ, Kunsch CA;
DR  WPI; 1999-045171/04.
XX
PT  New isolated Enterococcus faecalis polynucleotides and polypeptides
PT  - used to develop products for the detection of Enterococcus and for
PT  use in vaccines for prevention or attenuation of Enterococcus
PT  infection.
XX
PS  Claim 1; Page 429-438; 2084pp; English.
XX
CC  A computer readable medium has been developed which has recorded on it
CC  982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC  AAX12938 to AAX13919 represent these nucleotide sequences which are
CC  primary nucleotide sequences, also known as contigs. The computer-based
CC  system can identify fragments of the Enterococcus faecalis genome with
CC  commercial importance. The products can be used to detect the presence
CC  of Enterococcus faecalis in samples. They can also be used for
CC  diagnosing Enterococcal infection in an animal and monitoring
CC  progression of disease, and for identifying agents which can be used to
CC  modulate the growth or pathogenicity of Enterococcus faecalis, or
CC  another related organism, in vivo or in vitro. In particular the
CC  polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC  can be used in vaccines to prevent or attenuate an Enterococcal
CC  infection.
XX
SQ  Sequence 17087 BP; 4734 A; 3683 C; 3245 G; 5415 T; 10 other;
XX
XX
alignment_scores:
    Quality: 155.50      Length: 240
    Ratio: 1.341        Gaps: 9
    Percent Similarity: 48.333    Percent Identity: 27.083
alignment_block:
US-09-558-284-2 x AAX12984 ..
Align seg 1/1 to: AAX12984 from: 1 to: 17087
10  LeuPheAspLeuAspGlyThrIleValSerThrValAlaIaGluIy 26
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2773  ATTTTGATTTGGATGTGCTACTGACACCGCCCAATTCTATATCA 2822
26  salatrrpThrLysLeuGlyTyrGlyValAspProSerGluLeuP 43
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||
2823  AGCTTGAAGCGTTGCGGATTCGCTAGTATCCCATGACGAACCT 2872
43  helysHisSer.....
    ||:|||||:|||||:|||||:|||||:|||||:|||||:
2873  TCACACGAAACATTAAAGCATTAGTCGATGATTCCTTGACCGTATT 2922
47  .....HisGlyAlaArgThrGlnGluValLeuArgArgPheProIy 61
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||
2923  TTAGTCCATGCGCCACCGCGAA.....AATGCGTTTACTCCAGC 2960
61  sleuAspAspThrAspAsnLysGlyValLeuAlaLeuGluLysAspIleA 78
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||
2961  A.....GAAGAAAGACGCTTAGCTCAGCAAAAAAATGANC 2995
78  laHis.....SerTyrLeuAspThrValSerLeuIleProGly 90
    :||:|||||:|||||:|||||:|||||:|||||:|||||

```

```

2996  ACTACGTTCAATTATTAGAACAAATTAAACAGACAGATGTTTACCTGGT 3045
91  AlagLubnsleuLeuSerLeuAspValAspThrGlnThrGlnIySly 107
    :||:|||||:|||||:|||||:|||||:|||||:|||||:
3046  GTCGTGCCATTGCTA.....CAGCAACGCCAAGCAGC 3077
107  sleuProGluArgLysTrpAlaIleValThrSerGlySerProTyrLeuA 124
    :||:|||||:|||||:|||||:|||||:|||||:|||||
3078  TCACATTCCTCGCGCGCGCTTCGCTTCACAAAAACGCGCC..... 3120
124  laPheSerTrpPheGluThrIleLeuLysAsnValGlyLysProLysVal 140
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||
3121  .....CTGATTTGGAAAAATTAGCGCGTCCGCGCTAC 3153
141  PheIleThrGlyPheAsp.....ValLysAsnGlyLysProAspPr 154
    :||:|||||:|||||:|||||:|||||:|||||:|||||
3154  TTCGCCACCAATTGTGCGATCCGCACTGTTAGTAAAGCAAACTGATTC 3203
154  ogLugIyTyrSerArgAlaArgAspLeuLeuArgGlnAspLeuGlnLeuT 171
    :||:|||||:|||||:|||||:|||||:|||||:|||||
3204  TGAATCTTTTTCAGCCGCGCTGACACTATT..... 3234
171  hrGlyLysGlnAspLeuLysTyrValAlaPheGluAspAlaProValGly 187
    :||:|||||:|||||:|||||:|||||:|||||:|||||
3235  ..GGCGTGTACCGCAAAAGCCATTGCTTTGAGATGACACATCGAGC 3282
188  IleLysAlaGlyLysAlaMetGlyAlaIleThrValGlyIleThrSerSe 204
    :||:|||||:|||||:|||||:|||||:|||||:|||||
3283  ATTGACGCGTTGAAGCGCGCGCATCTATGCCGCTTGCGGCGCAG 3332
204  rTyrAspLysSerValLeuPheAspAlaGlyAlaAspTyrValAlaLysA 221
    :|||||:|||||:|||||:|||||:|||||:|||||
3333  CCAACCTTA.....ATCGAGCGCGACATGCACAGTTTCTG 3367
221  spleuThrGlnValSerVal 227
    :|||||:|||||:|||||:|||||:|||||:|||||
3368  AATGACTGAACCTAGCGCTT 3387
seq_name: /SIDSI/gcdata/geneseq/NA1996.DAT:AAAT42063
seq_documentation_block:
ID  AAT42063 standard; DNA; 1830121 BP.
XX
AC  AAT42063;
XX
DT  14-SEP-1999 (first entry)
XX
DE  Haemophilus influenzae complete genome sequence.
XX
KW  Genome; bacterium; Haemophilus influenzae; computer readable medium;
KW  expression modulating fragment; regulation; gene expression; vector;
KW  organism; open reading frame; ORF; ds.
XX
OS  Haemophilus influenzae.
XX
PN  W09633276-A1.
XX
PD  24-OCT-1996.
XX
PF  22-APR-1996; 96WO-US05320.
XX
PR  07-JUN-1995; 95US-0487429.
PR  21-APR-1995; 95US-0426787.
PR  07-JUN-1995; 95US-0476102.
XX
PA  (HUMA-) HUMAN GENOME SCI INC.
PI  (UYO ) UNTV JOHNS HOPKINS.
XX
AD  Adams MD, Fleischmann RD, Smith HO, Venter JC, White O;
DR  WPI; 1996-485782/48.
XX
PT  Haemophilus influenzae Rd genome recorded on computer readable

```


Align seg 1/1 to reverse of: AAX13135 from: 1 to: 1521

```

10 LeuheaSpLeuaspGlyThrIleValSerThrThrValAlaIaGlyLeu 26
   :::::::::::::::::::::
1497 ATTTTGATGATGATGGGTTGTTTGTATGACGAGTTGATTACTATAC 1448
26 sAlaTrpThrLysLeuCysTygIuTygIyValAspProSer...GluL 42
   :::::::::::::::::::::
1447 ATCTACCAAAAGTAGCATGATGGGTTTCCATATAGTAAAGAG 1398
42 eupheLysHisSerHisGlyAlaArgThrGluVal.....Leu 55
   :::::::::::::::::::::
1397 TCATTTAGACTATGTTGGATTTCGATGAGAGATACAAAGAAATAC 1348
56 ArgArpPhePheProLysLeuAspAspThrAspAsnLysGlyValLeuAl 72
   :::::::::::::::::::::
1347 CGACGATTATATGCA.....TCCTATGCTCATGACAC 1316
72 aLeuGluLysAspIleAlaHisSerTyLeuAspThr..... 84
   :::::::::::::::::::::
1315 AGTGGAGAAATTTATCCGCGTCTTATGATGATACCTGCAGAGTTTC 1266
85 .....ValSerLeuIleProGlyAlaGluAsnLeuLeuSer 97
   :::::::::::::::::::::
1265 GATCTGCTAACGTCCTTAAACACAGGGGTGGTAATTCCTGACTTT 1216
98 LeuAspValAspThrGluThrGluLysLeuProGluArgLysTrpAl 114
   :::::::::::::::::::::
1215 TTGAT.....GATCAAAATTCCTGCA..... 1192
114 aIleValThrSerGlySerProTyLeuAlaPheSerTrpPheGluThrI 131
   :::::::::::::::::::::
1191 .TTAGTGTCTCAAGTACGTCGTCGACCA.....ATTGAATGT 1152
131 leuLysAsnValGlyLysProLysValPhe.....IleThrGly 144
   :::::::::::::::::::::
1151 TATTAAACCATGCTGATTCAGAGATGTTTGTAGTATGTTTCAGCA 1102
145 PheAspValLysAsnGlyLysProAspProGluGlyTrsArgAlaAr 161
   :::::::::::::::::::::
1101 GAGGATGTTAAACGAGCAAAACGCGACCAAAATTTTCAAAAGCTCG 1052
161 gaSpLeuLeuArgGlnAspLeuGlnLeuThrGlyLysGlnAspLeuLysT 178
   :::::::::::::::::::::
1051 T.....CACTTTAGCAGACAGAACGCCGCAAAA 1023
178 yValValPheGluAspAlaProValGlyIleLysAlaGlyLysAlaMet 194
   :::::::::::::::::::::
1022 CCTTGATTTTCGAGACTCTTCATGAGGTTAAGTCGGCTCATAGTCG 973
195 GlyAlaIleThrValGlyIleThrSerSerTyAspLysSerValLeuPh 211
   :::::::::::::::::::::
972 GGCATTCCAGTGCATGATGATGATTTATGCAACCACTAATATTAT 923
211 eaSpAlaGlyAlaAspTyValValLysAspLeuThrGlnVal.SerVal 227
   :::::::::::::::::::::
922 CCAAGGAAACATTACATGCTTGGAGATTACATCAAGGCCACCAT 873
228 ValLysAsnAsnGluAsnGlyIleValIleGlnValAsnAspProLeu 243
   :::::::::::::::::::::
872 ATTTAAATATATTAAC.....AACACCTCTTTG 843

```

seq_name: /SIDS1/gcgdata/geneseq/NA2000.DAT: AAC45665

seq_documentation_block:

ID AAC45665 standard; DNA: 1057 BP.

AC AAC45665:

XX 18-OCT-2000 (first entry)

XX Arabidopsis thaliana DNA fragment SEQ ID NO: 47317.

```

XX Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX Arabidopsis thaliana.
XX EP1033405-A2.
PD 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
PF 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139482.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.

```

PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0144816.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.

PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores:

Quality: 136.00 Length: 221
Ratio: 1.183 Gaps: 7
Percent Similarity: 52.036 Percent Identity: 25.792

alignment_block:

US-09-558-284-2 x AAC45665 ..

Align seg 1/1 to: AAC45665 from: 1 to: 1057

10 LeuPheaspleuaspGlyThrIleValSerThrThrValAlaIaGlu 26
|||||.....
182 CTGTTGATGTTGATGACACACTCTGTGACATCCATCCACTTAT 231
26 sAlaTrpThrIysLeucSyrTyGluTyGly.....V 37
|||||.....
232 TGCCCTTCACAGACTGCTTCAAGAGATTGTTTAAACATGCTGCCAA 281
37 aIasProSerGluLeuPheIysHisSerHisGlyAlaArgThrGlnGlu 53
:|||||:.....
282 TCGATGAGAAATCTTGTGAGAACATGCTGGAACACAAATTCGAA 331
54 ValLeuArgArgPhePheProLysLeuAspThrAspAsnLysGlyVal 70
:|:.....
332 ATTGCTCTACTTCTTCTCCCT.....GATGATGTTTCAAGAGGTTAAA 375

```
70  LleuAlaLeuGluIlyAspIleAlaHisSerTyrlleuAspThrValSerL 87
    ::::: |||||::: ::::: :::::
376  ATCTGTGTATGAAAGAAAGACCTTTACCGCAAAATTGTGCAGACAGAA 425
    87  euIleProGlyAlaGluAsnLeuLeuSerLeuAspValAspThrGlu 103
    ::|||
    426  TAAACCA.....CTGATGCGCTTATATAA 451
    104  ThrGlnLysLysLeuProGluArg.....LysThrPalaIleValThrSe 118
    |||||::: |||||::: |||||:::
452  CTGACCAAAATGATGAGATGATGATGAAACGAGCTGCGGTACACAA 501
    118  rGlySerProTyrlleuAlaPheSerTrpPheGluThrIleLeuLysAsnV 135
    ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
502  CGCTCCAAAGAAACCGCA.....GAGCTCATGATATCAAC 539
    135  aGlyLysProLysValPhe.....IleThrGlyPheAspValLys 148
    ::||| ::||| ::||| ::||| ::|||
540  TTGGTCTGACTGATTCTTCAAGCAGTGTATCTTGCTCTGCAATGTCAA 589
    149  AsnGlyLysProAspProGluGluTyrlSerArgAlaArgAspLeuLar 165
    590  TTTCCCAACACACACCTGACCTTACTTGAAGGCTTGAAGTGCTTAA 639
    165  GlnAspLeuGlnLeuThrGlyLysGlnAspLeuLysTyrlValValPheG 182
    : ::::: ||| ::::: ||| ::::: ||| ::::: |||
640  CGTGCAAAGAGCACACG.....CTAGTTTCG 668
    182  LuAspAlaProValGlyIleLysAlaGlyLysAlaMetGlyAlaIleThr 198
    |||||::: |||||::: |||||:::
669  AAGACTCGATCTCTGGATMAAGCTGGAGTTGCACCTGCAATGCCAGTG 718
    199  ValGlyIleThrSerSerTyrlSerValIlePheAspAlaGlyAl 215
    ::|||::: |||||::: |||||:::
719  ATTGACACTACTACAGGGAATCCAGCAACCTGCTTATGCAGCAACAAACC 768
    215  AspTyrlValVal 219
    ::::: |||
769  CGCTTTTCTCATC 781

seq_name: /SIDS1/gcgdata/geneseq/geneseq/NR2000.DAT: AAC38884
seq_documentation_block:
ID  AAC38884 standard; DNA; 1060 BP.
XX
AC  AAC38884;
XX
DT  17-OCT-2000 (first entry)
XX
DE  Arabidopsis thaliana DNA fragment SEQ ID NO: 22573.
XX
KW  Hybridisation assay; genetic mapping; gene expression control;
KM  protein identification; signal transduction pathway;
KW  metabolic pathway; promoter; termination sequence; ss.
XX
OS  Arabidopsis thaliana.
XX
PN  EP1033405-A2.
XX
PD  06-SEP-2000.
XX
PF  25-FEB-2000; 2000EP-0301439.
XX
PR  25-FEB-1999; 99US-0121825.
PR  05-MAR-1999; 99US-0123180.
PR  09-MAR-1999; 99US-0123548.
PR  23-MAR-1999; 99US-0125788.
PR  25-MAR-1999; 99US-0126264.
PR  29-MAR-1999; 99US-0126785.
PR  01-APR-1999; 99US-0127462.
PR  06-APR-1999; 99US-0128234.
PR  08-APR-1999; 99US-0128714.
```

```
PR  16-APR-1999; 99US-0129845.
PR  19-APR-1999; 99US-0130077.
PR  21-APR-1999; 99US-0130049.
PR  23-APR-1999; 99US-0130510.
PR  23-APR-1999; 99US-0130891.
PR  28-APR-1999; 99US-0131449.
PR  30-APR-1999; 99US-0132048.
PR  30-APR-1999; 99US-0132407.
PR  04-MAY-1999; 99US-0132484.
PR  05-MAY-1999; 99US-0132485.
PR  06-MAY-1999; 99US-0132486.
PR  06-MAY-1999; 99US-0132487.
PR  07-MAY-1999; 99US-0132863.
PR  11-MAY-1999; 99US-0134256.
PR  14-MAY-1999; 99US-0134218.
PR  14-MAY-1999; 99US-0134219.
PR  14-MAY-1999; 99US-0134221.
PR  14-MAY-1999; 99US-0134370.
PR  18-MAY-1999; 99US-0134768.
PR  19-MAY-1999; 99US-0134941.
PR  20-MAY-1999; 99US-0135124.
PR  21-MAY-1999; 99US-0135353.
PR  24-MAY-1999; 99US-0135629.
PR  25-MAY-1999; 99US-0136021.
PR  27-MAY-1999; 99US-0136392.
PR  28-MAY-1999; 99US-0136782.
PR  01-JUN-1999; 99US-0137222.
PR  03-JUN-1999; 99US-0137528.
PR  04-JUN-1999; 99US-0137502.
PR  07-JUN-1999; 99US-0137724.
PR  08-JUN-1999; 99US-0138094.
PR  10-JUN-1999; 99US-0138540.
PR  10-JUN-1999; 99US-0138847.
PR  14-JUN-1999; 99US-0139119.
PR  16-JUN-1999; 99US-0139452.
PR  17-JUN-1999; 99US-0139453.
PR  18-JUN-1999; 99US-0139454.
PR  18-JUN-1999; 99US-0139455.
PR  18-JUN-1999; 99US-0139456.
PR  18-JUN-1999; 99US-0139457.
PR  18-JUN-1999; 99US-0139458.
PR  18-JUN-1999; 99US-0139459.
PR  18-JUN-1999; 99US-0139460.
PR  18-JUN-1999; 99US-0139461.
PR  18-JUN-1999; 99US-0139462.
PR  18-JUN-1999; 99US-0139463.
PR  18-JUN-1999; 99US-0139750.
PR  18-JUN-1999; 99US-0139763.
PR  21-JUN-1999; 99US-0139817.
PR  22-JUN-1999; 99US-0139899.
PR  23-JUN-1999; 99US-0140353.
PR  23-JUN-1999; 99US-0140354.
PR  24-JUN-1999; 99US-0140695.
PR  28-JUN-1999; 99US-0140823.
PR  29-JUN-1999; 99US-0140991.
PR  30-JUN-1999; 99US-0141287.
PR  01-JUL-1999; 99US-0141842.
PR  01-JUL-1999; 99US-0142154.
PR  02-JUL-1999; 99US-0142055.
PR  06-JUL-1999; 99US-0142390.
PR  08-JUL-1999; 99US-0142803.
PR  09-JUL-1999; 99US-0142920.
PR  12-JUL-1999; 99US-0142977.
PR  13-JUL-1999; 99US-0143542.
PR  14-JUL-1999; 99US-0143624.
PR  15-JUL-1999; 99US-0144005.
PR  16-JUL-1999; 99US-0144085.
PR  16-JUL-1999; 99US-0144086.
PR  19-JUL-1999; 99US-0144325.
PR  19-JUL-1999; 99US-0144331.
PR  19-JUL-1999; 99US-0144332.
PR  19-JUL-1999; 99US-0144333.
```



```

590 TTTCCAAACACACCCCTGACCTTACTTGAAGCCTCTGAAGTCTTA 639
      |||||  |||  |||  ::|||  ::::|||:
165 gGlnAspLeuGlnLeuThrGlyLysGlnAspLeuLysTyrValValPheG 182
      ::::  ::  |||  ::||||||
640 CGTGTCAAGGAGCACACG.....CTAGTTTTCG 668
      |||||  |||||  |||||  |||  |||  |||
182 LAspAlaProValGlyIleLysAlaGlyLysAlaMetGlyAlaIleThr 198
      |||||  |||||  |||||  |||||  |||  |||
669 AAGACTCGATCTCTGGGATAAAGCTGAGTTGCAGCTGSAATGCCAGTG 718
      ::|||  ::|||  ::|||  ::|||  ::|||
199 ValGlyIleThrSerSerTyrAspLysSerValLeuPheAspAlaGlyAl 215
      ::|||  ::|||  ::|||  ::|||  ::|||
719 ATGGAGCTGACTACAGGAATCCAGCAAGCCCTTATGCACAAACAAAC 768
      ::|||  ::|||  ::|||  ::|||  ::|||
215 aAspTyrValVal 219
      ::|||  ::|||  ::|||  ::|||  ::|||
769 CGCTTTCTCATC 781

```


OM of: US-09-558-284-2 to: Issued_Patents_NA:* out_format : pfs

Date: Jul 22, 2001 9:28 PM

About: Results were produced by the GenCore software, version 4.5,

Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp
-O=/cgn2_1/USPTO.spool/US09558284/runat.20072001.160145_19160/app-query.fasta_1.306
-DB=Issued_Patents_NA -ORFMT=fastap -SUFFIX=p2n.rn1
-GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPECL=0.000
-LOOPEXT=0.000 -OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000
-XGAPEXT=0.500 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000
-YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1
-MATRIX=blsnum62 -TRANS=human40.cdi -LIST=45 -DOCALLIGN=200
-THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=200000000
-USER=US09558284@cgn1_1.60 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPHY
-WAIT -THREADS=1

Search information block:

Query: US-09-558-284-2
Query length: 246
Database: Issued_Patents_NA:*
Database sequences: 317530
Database length: 92630169
Search time (sec): 62.610000

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation	816	816	816	816
/cgn2_6/prodata/2/ina/6A.COMB.seq:US-08-968-563-8 +		328.50	751.09	3.1e-34	816	1	816	1	816
/cgn2_6/prodata/2/ina/6B.COMB.seq:US-08-969-683A-8 +		328.50	751.09	3.1e-34	816	1	816	1	816
/cgn2_6/prodata/2/ina/6A.COMB.seq:US-08-968-563-9 +		327.50	749.69	3.8e-34	753	1	753	1	753
/cgn2_6/prodata/2/ina/6B.COMB.seq:US-08-969-683A-9 +		327.50	749.69	3.8e-34	753	1	753	1	753
/cgn2_6/prodata/2/ina/6B.COMB.seq:US-08-968-416-874 +		118.00	253.74	1.6e-06	590	1	590	1	590
/cgn2_6/prodata/2/ina/5A.COMB.seq:US-08-363-311-14 +		109.50	223.11	8.0e-05	1380	1	1380	1	1380
/cgn2_6/prodata/2/ina/5A.COMB.seq:US-08-463-288A-14 +		109.50	223.11	8.0e-05	1380	1	1380	1	1380
/cgn2_6/prodata/2/ina/5B.COMB.seq:US-08-470-445A-14 +		109.50	223.11	8.0e-05	1380	1	1380	1	1380
/cgn2_6/prodata/2/ina/5B.COMB.seq:US-08-462-679-14 +		109.50	223.11	8.0e-05	1380	1	1380	1	1380
/cgn2_6/prodata/2/ina/5B.COMB.seq:US-08-466-210A-14 +		109.50	223.11	8.0e-05	1380	1	1380	1	1380
/cgn2_6/prodata/2/ina/5B.COMB.seq:US-08-467-147A-14 +		109.50	223.11	8.0e-05	1380	1	1380	1	1380
/cgn2_6/prodata/2/ina/5B.COMB.seq:US-08-469-014-14 +		109.50	223.11	8.0e-05	1380	1	1380	1	1380
/cgn2_6/prodata/2/ina/5B.COMB.seq:US-08-904-263A-3 +		82.50	146.34	1.52	3825	1	3825	1	3825
/cgn2_6/prodata/2/ina/5A.COMB.seq:US-08-119-125A-3 +		82.50	145.44	1.71	4118	1	4118	1	4118
/cgn2_6/prodata/2/ina/6B.COMB.seq:US-09-433-262-1 +		82.50	130.40	11.73	14078	1	14078	1	14078
/cgn2_6/prodata/2/ina/6B.COMB.seq:US-08-859-694-1 +		82.50	129.98	12.33	14578	1	14578	1	14578
/cgn2_6/prodata/2/ina/5B.COMB.seq:US-08-892-770-2 +		80.50	149.95	0.9559	1929	1	1929	1	1929
/cgn2_6/prodata/2/ina/6A.COMB.seq:US-09-162-373-2 +		79.50	138.86	3.96	3930	1	3930	1	3930
/cgn2_6/prodata/2/ina/5B.COMB.seq:US-08-710-584-3 +		86.50	164.40	0.1497	1903	1	1903	1	1903
/cgn2_6/prodata/2/ina/6A.COMB.seq:US-08-904-263A-3 +		82.50	146.34	1.52	3825	1	3825	1	3825
/cgn2_6/prodata/2/ina/5A.COMB.seq:US-08-119-125A-3 +		82.50	145.44	1.71	4118	1	4118	1	4118
/cgn2_6/prodata/2/ina/6B.COMB.seq:US-09-433-262-1 +		82.50	130.40	11.73	14078	1	14078	1	14078
/cgn2_6/prodata/2/ina/6B.COMB.seq:US-08-859-694-1 +		82.50	129.98	12.33	14578	1	14578	1	14578
/cgn2_6/prodata/2/ina/5B.COMB.seq:US-08-892-770-2 +		80.50	149.95	0.9559	1929	1	1929	1	1929
/cgn2_6/prodata/2/ina/6A.COMB.seq:US-09-162-373-2 +		79.50	138.86	3.96	3930	1	3930	1	3930
/cgn2_6/prodata/2/ina/5B.COMB.seq:US-08-710-584-3 +		86.50	164.40	0.1497	1903	1	1903	1	1903
/cgn2_6/prodata/2/ina/6A.COMB.seq:US-08-904-263A-3 +		82.50	146.34	1.52	3825	1	3825	1	3825
/cgn2_6/prodata/2/ina/5A.COMB.seq:US-08-119-125A-3 +		82.50	145.44	1.71	4118	1	4118	1	4118
/cgn2_6/prodata/2/ina/6B.COMB.seq:US-09-433-262-1 +		82.50	130.40	11.73	14078	1	14078	1	14078
/cgn2_6/prodata/2/ina/6B.COMB.seq:US-08-859-694-1 +		82.50	129.98	12.33	14578	1	14578	1	14578
/cgn2_6/prodata/2/ina/5B.COMB.seq:US-08-892-770-2 +		80.50	149.95	0.9559	1929	1	1929	1	1929
/cgn2_6/prodata/2/ina/6A.COMB.seq:US-09-162-373-2 +		79.50	138.86	3.96	3930	1	3930	1	3930
/cgn2_6/prodata/2/ina/5B.COMB.seq:US-08-710-584-3 +		86.50	164.40	0.1497	1903	1	1903	1	1903
/cgn2_6/prodata/2/ina/6A.COMB.seq:US-08-904-263A-3 +		82.50	146.34	1.52	3825	1	3825	1	3825
/cgn2_6/prodata/2/ina/5A.COMB.seq:US-08-119-125A-3 +		82.50	145.44	1.71	4118	1	4118	1	4118
/cgn2_6/prodata/2/ina/6B.COMB.seq:US-09-433-262-1 +		82.50	130.40	11.73	14078	1	14078	1	14078
/cgn2_6/prodata/2/ina/6B.COMB.seq:US-08-859-694-1 +		82.50	129.98	12.33	14578	1	14578	1	14578
/cgn2_6/prodata/2/ina/5B.COMB.seq:US-08-892-770-2 +		80.50	149.95	0.9559	1929	1	1929	1	1929
/cgn2_6/prodata/2/ina/6A.COMB.seq:US-09-162-373-2 +		79.50	138.86	3.96	3930	1	3930	1	3930
/cgn2_6/prodata/2/ina/5B.COMB.seq:US-08-710-584-3 +		86.50	164.40	0.1497	1903	1	1903	1	1903
/cgn2_6/prodata/2/ina/6A.COMB.seq:US-08-904-263A-3 +		82.50	146.34	1.52	3825	1	3825	1	3825
/cgn2_6/prodata/2/ina/5A.COMB.seq:US-08-119-125A-3 +		82.50	145.44	1.71	4118	1	4118	1	4118
/cgn2_6/prodata/2/ina/6B.COMB.seq:US-09-433-262-1 +		82.50	130.40	11.73	14078	1	14078	1	14078
/cgn2_6/prodata/2/ina/6B.COMB.seq:US-08-859-694-1 +		82.50	129.98	12.33	14578	1	14578	1	14578
/cgn2_6/prodata/2/ina/5B.COMB.seq:US-08-892-770-2 +		80.50	149.95	0.9559	1929	1	1929	1	1929
/cgn2_6/prodata/2/ina/6A.COMB.seq:US-09-162-373-2 +		79.50	138.86	3.96	3930	1	3930	1	3930
/cgn2_6/prodata/2/ina/5B.COMB.seq:US-08-710-584-3 +		86.50	164.40	0.1497	1903	1	1903	1	1903
/cgn2_6/prodata/2/ina/6A.COMB.seq:US-08-904-263A-3 +		82.50	146.34	1.52	3825	1	3825	1	3825
/cgn2_6/prodata/2/ina/5A.COMB.seq:US-08-119-125A-3 +		82.50	145.44	1.71	4118	1	4118	1	4118
/cgn2_6/prodata/2/ina/6B.COMB.seq:US-09-433-262-1 +		82.50	130.40	11.73	14078	1	14078	1	14078
/cgn2_6/prodata/2/ina/6B.COMB.seq:US-08-859-694-1 +		82.50	129.98	12.33	14578	1	14578	1	14578
/cgn2_6/prodata/2/ina/5B.COMB.seq:US-08-892-770-2 +		80.50	149.95	0.9559	1929	1	1929	1	1929
/cgn2_6/prodata/2/ina/6A.COMB.seq:US-09-162-373-2 +		79.50	138.86	3.96	3930	1	3930	1	3930
/cgn2_6/prodata/2/ina/5B.COMB.seq:US-08-710-584-3 +		86.50	164.40	0.1497	1903	1	1903	1	1903
/cgn2_6/prodata/2/ina/6A.COMB.seq:US-08-904-263A-3 +		82.50	146.34	1.52	3825	1	3825	1	3825
/cgn2_6/prodata/2/ina/5A.COMB.seq:US-08-119-125A-3 +		82.50	145.44	1.71	4118	1	4118	1	4118
/cgn2_6/prodata/2/ina/6B.COMB.seq:US-09-433-262-1 +		82.50	130.40	11.73	14078	1	14078	1	14078
/cgn2_6/prodata/2/ina/6B.COMB.seq:US-08-859-694-1 +		82.50	129.98	12.33	14578	1	14578	1	14578
/cgn2_6/prodata/2/ina/5B.COMB.seq:US-08-892-770-2 +		80.50	149.95	0.9559	1929	1	1929	1	1929
/cgn2_6/prodata/2/ina/6A.COMB.seq:US-09-162-373-2 +		79.50	138.86	3.96	3930	1	3930	1	3930
/cgn2_6/prodata/2/ina/5B.COMB.seq:US-08-710-584-3 +		86.50	164.40	0.1497	1903	1	1903	1	1903
/cgn2_6/prodata/2/ina/6A.COMB.seq:US-08-904-263A-3 +		82.50	146.34	1.52	3825	1	3825	1	3825
/cgn2_6/prodata/2/ina/5A.COMB.seq:US-08-119-125A-3 +		82.50	145.44	1.71	4118	1	4118	1	4118
/cgn2_6/prodata/2/ina/6B.COMB.seq:US-09-433-262-1 +		82.50	130.40	11.73	14078	1	14078	1	14078
/cgn2_6/prodata/2/ina/6B.COMB.seq:US-08-859-694-1 +		82.50	129.98	12.33	14578	1	14578	1	14578
/cgn2_6/prodata/2/ina/5B.COMB.seq:US-08-892-770-2 +		80.50	149.95	0.9559	1929	1	1929	1	1929
/cgn2_6/prodata/2/ina/6A.COMB.seq:US-09-162-373-2 +		79.50	138.86	3.96	3930	1	3930	1	3930
/cgn2_6/prodata/2/ina/5B.COMB.seq:US-08-710-584-3 +		86.50	164.40	0.1497	1903	1	1903	1	1903
/cgn2_6/prodata/2/ina/6A.COMB.seq:US-08-904-263A-3 +		82.50	146.34	1.52	3825	1	3825	1	3825
/cgn2_6/prodata/2/ina/5A.COMB.seq:US-08-119-125A-3 +		82.50	145.44	1.71	4118	1	4118	1	4118
/cgn2_6/prodata/2/ina/6B.COMB.seq:US-09-433-262-1 +		82.50	130.40	11.73	14078	1	14078	1	14078
/cgn2_6/prodata/2/ina/6B.COMB.seq:US-08-859-694-1 +		82.50	129.98	12.33	14578	1	14578	1	14578
/cgn2_6/prodata/2/ina/5B.COMB.seq:US-08-892-770-2 +		80.50	149.95	0.9559	1929	1	1929	1	1929
/cgn2_6/prodata/2/ina/6A.COMB.seq:US-09-162-373-2 +		79.50	138.86	3.96	3930	1	3930	1	3930
/cgn2_6/prodata/2/ina/5B.COMB.seq:US-08-710-584-3 +		86.50	164.40	0.1497	1903	1	1903	1	1903
/cgn2_6/prodata/2/ina/6A.COMB.seq:US-08-904-263A-3 +		82.50	146.34	1.52	3825	1	3825	1	3825
/cgn2_6/prodata/2/ina/5A.COMB.seq:US-08-119-125A-3 +		82.50	145.44	1.71	4118	1	4118	1	4118
/cgn2_6/prodata/2/ina/6B.COMB.seq:US-09-433-262-1 +		82.50	130.40	11.73	14078	1	14078	1	14078
/cgn2_6/prodata/2/ina/6B.COMB.seq:US-08-859-694-1 +		82.50	129.98	12.33	14578	1	14578	1	14578
/cgn2_6/prodata/2/ina/5B.COMB.seq:US-08-892-770-2 +		80.50	149.95	0.9559	1929	1	1929	1	1929
/cgn2_6/prodata/2/ina/6A.COMB.seq:US-09-162-373-2 +		79.50	138.86	3.96	3930	1	3930	1	3930
/cgn2_6/prodata/2/ina/5B.COMB.seq:US-08-710-584-3 +		86.50	164.40	0.1497	1903	1	1903	1	1903
/cgn2_6/prodata/2/ina/6A.COMB.seq:US-08-904-263A-3 +		82.50	146.34	1.52	3825	1	3825	1	3825
/cgn2_6/prodata/2/ina/5A.COMB.seq:US-08-119-125A-3 +		82.50	145.44	1.71	4118	1	4118	1	4118
/cgn2_6/prodata/2/ina/6B.COMB.seq:US-09-433-262-1 +		82.50	130.40	11.73	14078	1	14078	1	14078
/cgn2_6/prodata/2/ina/6B.COMB.seq:US-08-859-694-1 +		82.50	129.98	12.33	14578	1	14578	1	14578
/cgn2_6/prodata/2/ina/5B.COMB.seq:US-08-892-770-2 +		80.50	149.95	0.9559	1929	1	1929	1	1929
/cgn2_6/prodata/2/ina/6A.COMB.seq:US-09-162-373-2 +		79.50	138.86	3.96	3930	1	3930	1	3930
/cgn2_6/prodata/2/ina/5B.COMB.seq:US-08-710-584-3 +		86.50	164.40	0.1497	1903	1	1903	1	1903
/cgn2_6/prodata/2/ina/6A.COMB.seq:US-08-904-263A-3 +		82.50	146.34	1.52	3825	1	3825	1	3825
/cgn2_6/prodata/2/ina/5A.COMB.seq:US-08-119-125A-3 +		82.50	145.44	1.71	4118	1	4118	1	4118
/cgn2_6/prodata/2/ina/6B.COMB.seq:US-09-433-262-1 +		82.50	130.40	11.73	14078	1	14078	1	14078
/cgn2_6/prodata/2/ina/6B.COMB.seq:US-08-859-694-1 +		82.50	129.98	12.33	14578	1	14578	1	14578
/cgn2_6/prodata/2/ina/5B.COMB.seq:US-08-892-770-2 +		80.50	149.95	0.9559	1929	1	1929	1	1929
/cgn2_6/prodata/2/ina/6A.COMB.seq:US-09-162-373-2 +		79.50	138.86	3.96	3930	1	3930	1	3930
/cgn2_6/prodata/2/ina/5B.COMB.seq:US-08-710-584-3 +		86.50	164.40	0.1497	1903	1	1903	1	1903
/cgn2_6/pro									

```

: GENERAL INFORMATION:
: APPLICANT: GENENCOR INTERNATIONAL, INC.
: TITLE OF INVENTION: METHOD FOR THE RECOMBINANT
: TITLE OF INVENTION: PRODUCTION OF 1,3 PROPEANEDIO

```

[illegible]

291 CAAGCTAGAGGTGAAATCCAGAAAGTACGGTGAACACTCCATCGAAG 340

OTHER INFORMATION: may be repeated up to eight times (for a total of
 OTHER INFORMATION: nine repeating copies of these sequences within
 OTHER INFORMATION: the polynucleotide)."

US-08-363-311-14

alignment_scores: Length: 229
 Quality: 109.50
 Ratio: 0.928
 Percent Similarity: 51.528 Percent Identity: 24.891

alignment_block:

US-09-558-284-2 x US-08-363-311-14 ..

Align seg 1/1 to: US-08-363-311-14 from: 1 to: 1380

```

27  ALaTrpThrLysLeuCySerGlu...TyrGlyValAspProSerGluLeu 42
   |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
31  GCCTTAATTTACTCTGTATAGAGTTATACAGATTAAGAGAAATATTAT 80
   |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
42  uphellyshisser.....HisGlyAlaArgThrGlnGluValLeuArg 57
   |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
81  GTTAGAAGAGCTTAATAAATACAGTTATGACTTCACAGAGAAACAC 130
   |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
57  rgpherphepTolysLeuAspSprThrAspAsnLysGlyValLeuAla 73
   |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
131 GGTTCCTCAATTAG..... 144
   |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
74  GluLysAspIleAlaHisSerTyrLeuAspThrValSerLeuIlePro 90
   |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
145 .....AGTTCAAGTTGG 158
   |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
90  yAlaGluAsnLeuLeuSerLeuAsp...ValAspThrGlnThrGln 106
   |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
159  TGCACCTCTGTACTAATGCTGTAGTTTTCGCGGCTTACACAG 208
   |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
106  yslLysLeuProGluArgLysThrAlaIleValThrSerGlySer...Pro 121
   |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
209  GTAATCTTAATATTTTGAAGAGTCAATATGTCGCATCTCAATTC 258
   |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
122  TyrLeuAlaPheSerTyrPheGluThrIleLeuLysAsnValGlyLys 138
   |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
259  GCGAGTCGACGACCTTAATACAGCATCACTAAATAATACAAACG 308
   |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
138  oLysValPheIleThrGlyPheAspValLysAsnGlyLysProAspPro 154
   |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
309  AAATGCTTACATAGATTATATGATGATAAATTAGTAAATAGATCCAT 358
   |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
155 .....GluGlyTyrSerArgAlaArgAspLeu 164
   |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
359  TACATTAATTTTATAGAACAGGTTTACA...GCAAGATATGTTTTT 405
   |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
165  ArgGln.....AspLeuGlnLeuThrGln 172
   |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
406  AGACAAGTACTAATAATACTATGCGGATGTTCTCAGTTTCACAGTAC 455
   |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
172  yLysGlnAspLeuLysTyrValValPhe...GluAspAlaProValGly 188
   |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
456  AAGGCTACTCTTACCTAATATATTTGGTGAAGATGACATACCCATG 505
   |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
188  LeLysAlaGlyLysAlaMetGlyAlaIleThrValGlyIleThrSer 204
   |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
506  TAAAGCTCATGACCAAAATGATATAGTATGTTGCTCTTAAT...ATT 552
   |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
205  TyrAspLysSerValLeuPheAspAlaGlyAlaAspTyrValVal 221
   |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
553  TATGATTCACAACCTTGTGAGGAT..... 576
   |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
221  pleuThrGlnValSerValValLysAsnGluAsn 233
   |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
577 .....AAGATTGAAGAACTTAGAACATGCAAC 606
   |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

seq_name: /cgn2_6/ptodata/2/lna/5A_COMB.seq:US-08-463-288A-14

seq documentation_block:

Sequence 14, Application US/08463288A

Patent No. 5820860

GENERAL INFORMATION:

APPLICANT: Michel, James L.

APPLICANT: Kasper, Dennis L.

APPLICANT: Ausubel, Frederick M.

APPLICANT: Madoff, Lawrence C.

TITLE OF INVENTION: Conjugate Vaccine For Group B

TITLE OF INVENTION: Streptococcus

NUMBER OF SEQUENCES: 65

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.

STREET: 1100 New York Avenue, NW, Suite 600

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20005-3934

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/463,288A

FILING DATE: 05-JUN-1995

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/363,311

FILING DATE: 22-DEC-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/968,866

FILING DATE: 02-NOV-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/408,036

FILING DATE: 15-SEP-1989

ATTORNEY/AGENT INFORMATION:

NAME: Bugalsky, Lawrence B.

REGISTRATION NUMBER: 35,086

REFERENCE/DOCKET NUMBER: 0609,2370007

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 371-2600

TELEFAX: (202) 371-2540

TELEX: 248636 SSK

INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:

LENGTH: 1380 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

FEATURE:

NAME/KEY: CDS

LOCATION: 79..1173

FEATURE:

NAME/KEY: misc-feature

LOCATION: 1004

OTHER INFORMATION:

OTHER INFORMATION:

OTHER INFORMATION:

OTHER INFORMATION:

OTHER INFORMATION:

OTHER INFORMATION:

OTHER INFORMATION:

OTHER INFORMATION:

OTHER INFORMATION:

OTHER INFORMATION:

OTHER INFORMATION:

OTHER INFORMATION:

OTHER INFORMATION:

OTHER INFORMATION:

OTHER INFORMATION:

OTHER INFORMATION:

OTHER INFORMATION:

OTHER INFORMATION:

OTHER INFORMATION:

OTHER INFORMATION:

OTHER INFORMATION:

OTHER INFORMATION:

OTHER INFORMATION:

alignment_scores:

Quality: 109.50

Ratio: 0.928

Percent Similarity: 51.528

Percent Identity: 24.891

alignment_block:

US-09-558-284-2 x US-08-463-288A-14 ..

seq_name: /cgn2_6/ptodata/2/lna/5A_COMB.seq:US-08-463-288A-14

Align seq 1/1 to: US-08-463-288A-14 from: 1 to: 1380

```
27 AlatrphrlysluGcysTyrGlu...TyrGlyValAspProSerGluLe 42
|||||:||||| ||| ||| |||:
31 GCCATATATTACTCTGTATAGACTTATACAGACTAAAGGAGATATTAT 80
|||||:||||| ||| ||| |||:
42 uphelyshisSer.....HisGlyAlaArgThrGlnGluValLeuArg 57
|||||:||||| ||| ||| |||:
81 GTTAGAAGGCTAAATAATACACTTATGATACCTTCACAGCAACAAAC 130
|||||:||||| ||| ||| |||:
57 rghpehpelolysLeuAspAspThrAspAsnLysGlyValLeuAlaLeu 73
|||||:||||| ||| ||| |||:
131 GGTTCATTAATTAG..... 144
|||||:||||| ||| ||| |||:
74 GluLysAspLeuAlaHisSerTyrLeuAspThrValSerLeuIleProG 90
|||||:||||| ||| ||| |||:
145 .....AAGTTCAAGCTTGG 158
|||||:||||| ||| ||| |||:
90 yAlaGluAsnLeuLeuSerLeuAsp...ValAspThrGlnThrGlnL 106
|||||:||||| ||| ||| |||:
159 TCGAGCTTCTGTACTATGTGGTCTTACTTTTGGTGGGCTTACACAG 208
|||||:||||| ||| ||| |||:
106 yslLysLeuProGluArgLysTrpAlaIleValThrSerGlySer... 121
|||||:||||| ||| ||| |||:
209 CTATCTTATATTATTTTGAAGAGCTCAATAGTCTGCATCTACAATT 258
|||||:||||| ||| ||| |||:
122 TyrLeuAlaPheSerTrpPheGluThrIleLeuLysAsnValGlyLys 138
|||||:||||| ||| ||| |||:
259 GCGAGCTCAGCAGCTTAATACAGACATCAGTAAATATACAAACGCG 308
|||||:||||| ||| ||| |||:
138 olYsValPheIleThrGlyPheAspValLysAsnGlyLysProAspPro 154
|||||:||||| ||| ||| |||:
309 AAATGCTTACATGATTTATATGATGATAAATTAGTAAATATGATCCAT 358
|||||:||||| ||| ||| |||:
155 .....GluGlyTyrSerArgAlaArgAspLeu 164
|||||:||||| ||| ||| |||:
359 TACAATTAATGTTTGTAGAACAGCTTTTACA...GCCAAGTATGTTT 405
|||||:||||| ||| ||| |||:
165 ArgGln.....AspLeuGlnLeuThrGln 172
|||||:||||| ||| ||| |||:
406 AGCACAAGCTACTAATCTATGGGAGCTTCTCAGTTGCCAGACTACAGG 455
|||||:||||| ||| ||| |||:
172 yLysGlnAspLeuLysTyrValValPhe...GluAspAlaProValGly 188
|||||:||||| ||| ||| |||:
456 AAGCGTAGTACTTACCTATATATATTTGGTGAAGATGACTACCATG 505
|||||:||||| ||| ||| |||:
188 lelysAlaGlyLysAlaMetGlyAlaIleThrValGlyIleThrSerSer 204
|||||:||||| ||| ||| |||:
506 TAAAGACTGATGCAACAATGATGATAGTGTGCTTAACT...ATT 552
|||||:||||| ||| ||| |||:
205 TyrAspLysSerValLeuPheAspAlaGlyAlaAspTyrValValLys 221
|||||:||||| ||| ||| |||:
553 TATGATTCACAACCTTGAGCGAT..... 576
|||||:||||| ||| ||| |||:
221 PleuThrGlnValSerValValLysAsnAsnGluAsn 233
|||||:||||| ||| ||| |||:
577 .....AAGATTGAAGAGATTGACAGCAATGCCAAC 606
|||||:||||| ||| ||| |||:
```

seq_name: /cgn2.6/plodata/2/ina/5B_COMB.seq:US-08-470-445A-14

seq_documentation_block:

Sequence 14, Application US/0847045A

Patent No. 5833444

GENERAL INFORMATION:

APPLICANT: Michel, James L.

APPLICANT: Kasper, Dennis L.

APPLICANT: Ausubel, Frederick M.

APPLICANT: Madoff, Lawrence C.

TITLE OF INVENTION: Conjugate Vaccine Against Group B

NUMBER OF SEQUENCES: 65

CORRESPONDENCE ADDRESS:

```
ADDRESS: Sterne, Kessler, Goldstein & Fox P.L.L.C.
STREET: 1100 New York Avenue, NW, Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/470,445A
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/363,311
FILING DATE: 22-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/968,866
FILING DATE: 02-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/408,036
FILING DATE: 15-SEP-1989
ATTORNEY/AGENT INFORMATION:
NAME: Bugalsky, Lawrence B.
REGISTRATION NUMBER: 35,086
REFERENCE/DOCKET NUMBER: 0609.237000A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
TELEX: 248636 SSK
INFORMATION FOR SEO ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 1380 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 79..1173
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1004
OTHER INFORMATION: /note="this feature is to signify
OTHER INFORMATION: that the nucleotide sequence from position 757
OTHER INFORMATION: through 1003 is inserted at position 1004 and can be
OTHER INFORMATION: repeated up to eight times (for a total of nine
OTHER INFORMATION: other INFORMATION: repeating copies of these sequences within the
US-08-470-445A-14
alignment_scores:
Quality: 109.50 Length: 229
Ratio: 0.928 Gaps: 11
Percent Similarity: 51.528 Percent Identity: 24.891
alignment_block:
US-09-558-284-2 x US-08-470-445A-14 ..
Align seq 1/1 to: US-08-470-445A-14 from: 1 to: 1380
27 AlatrphrlysluGcysTyrGlu...TyrGlyValAspProSerGluLe 42
|||||:||||| ||| ||| |||:
31 GCCATATATTACTCTGTATAGACTTATACAGACTAAAGGAGATATTAT 80
|||||:||||| ||| ||| |||:
42 uphelyshisSer.....HisGlyAlaArgThrGlnGluValLeuArg 57
|||||:||||| ||| ||| |||:
81 GTTAGAAGGCTAAATAATACACTTATGATACCTTCACAGCAACAAAC 130
|||||:||||| ||| ||| |||:
57 rghpehpelolysLeuAspAspThrAspAsnLysGlyValLeuAlaLeu 73
|||||:||||| ||| ||| |||:
```

```

131 GCTTTCAATTAG..... 144
74 GluLysAspIleAlaHisSerTyrLeuAspThrValSerLeuIleProG1 90
145 .....AGTTCAAGTTTGG 158
90 yAlaGluAsnLeuLeuSerLeuAsp...ValAspThrGluThrGlnL 106
159 TGCAGCTTGCTACTAATTTGGTTAGTTTGGTGGGCTTACACAG 208
106 yLysLeuProGluArgLysTrpAlaIleValThrSerGlySer...Pro 121
209 GTATCTTAATATTTTGAAGACATAGTTGGCATCTACAAATTCGA 258
122 TyrLeuAlaPheSerTrpPheGluThrIleLeuLysAsnValGlyLysPr 138
259 GGCAGCTGCAGCGACCTTAATACAGCATCAATAAATATACAAACGG 308
138 oLysValPheIleThrGlyPheAspValLysAsnGlyLysProAspPro 154
309 AAATGCTTACATAGATTATATGATGTAATAATTAGCTAAATAGATCCAT 358
155 .....GluGlyTyrSerArgAlaArgAspLeuLeu 164
359 TACAATTATTTGTTTACACACAGGTTTACA...GCAAGATATGTTT 405
165 ArgGln.....AspLeuGlnLeuThrG1 172
406 AGACAAAGTACTAAATACTATGCGGATGTTTCTCAGTGCACAGATCAGG 455
172 yLysGlnAspLeuLysTyrValIleValPhe...GluAspAlaProValGlyI 188
456 AAGGCTGCTCTTACCTTAATATATTTGGTCAACATGACATCACACATG 505
188 lLysAlaGlyLysAlaMetGlyAlaIleThrValGlyIleThrSerSer 204
506 TAAAGACTGATGACAAATTCATATAGTAGTTCGCTTAACT...AAT 552
205 TyrAspLysSerValLeuPheAspAlaGlyAlaAspTyrValIleCysAs 221
553 TATGATTCACACACCTTGAGGAT..... 576
221 PLeuThrGlnValSerValIleValLysAsnGlnLysn 233
577 .....AAGATTGAAGAAGTTAGAACGAAATGCAAC 606

seq.name: /cgn2_6/ptodata/2/ina/58_COMB.seq:US-08-462-679-14

seq-documentation_block:
; Sequence 14, Application US/08462679
; Patent No. 5847081
; GENERAL INFORMATION:
; APPLICANT: Michel, James L.
; APPLICANT: Kasper, Dennis L.
; APPLICANT: Ausubel, Frederick M.
; APPLICANT: Madoff, Lawrence C.
; TITLE OF INVENTION: Conjugate Vaccine For Group B
; NUMBER OF SEQUENCES: 65
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, NW, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/462.679

```

```

; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/363,311
; FILING DATE: 22-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/968,866
; FILING DATE: 02-NOV-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/408,036
; FILING DATE: 15-SEP-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Bugalsky, Lawrence B.
; REGISTRATION NUMBER: 35,086
; REFERENCE/DOCKET NUMBER: 0609.2370008
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; TELEX: 248636 SSK
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1380 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 79..1173
; NAME/KEY: misc_feature
; LOCATION: 1004
; OTHER INFORMATION: /note="This feature is to signify
; OTHER INFORMATION: that the nucleotide sequence from position 757
; OTHER INFORMATION: through 1003 is inserted at position 1004 and can be
; OTHER INFORMATION: repeated up to eight times (for a total of nine
; OTHER INFORMATION: repeating copies of these sequences within the
; US-08-462-679-14

alignment_scores:
Quality: 109.50 Length: 229
Ratio: 0.928 Gaps: 11
Percent Similarity: 51.528 Percent Identity: 24.891

alignment_block:
US-09-558-284-2 x US-08-462-679-14 ..
Align seg 1/1 to: US-08-462-679-14 from: 1 to: 1380

27 AlATrPThLysLeuCySTyrGlu...TyrGlyValAlAspProSerG1Le 42
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 80
31 GCCTAATATTACTGTATAGATGATACAGATTAAGGAGAAATATATAT 80
42 uPheLysHisSer.....HisGlyValArgThrGlnGluValLeuArg 57
|||||:|||||:|||||:|||||:|||||:|||||:|||||: 130
81 GTTAAAGGCTCAAAAATACAGTTATGATGATCTTACAGACAAACAC 130
57 rGPhePheProLysLeuAspAspThrAspAsnLysGlyValLeuAlaLeu 73
|||||:|||||:|||||:|||||:|||||:|||||:|||||: 144
131 GCTTTCAATTAG.....AGTTCAAGTTTGG 158
74 GluLysAspIleAlaHisSerTyrLeuAspThrValSerLeuIleProG1 90
145 .....AGTTCAAGTTTGG 158
90 yAlaGluAsnLeuLeuSerLeuAsp...ValAspThrGluThrGlnL 106
159 TGCAGCTTGCTACTAATTTGGTTAGTTTGGTGGGCTTACACAG 208
106 yLysLeuProGluArgLysTrpAlaIleValThrSerGlySer...Pro 121
209 GTATCTTAATATTTTGAAGACATAGTTGGCATCTACAAATTCGA 258

```



```

122 TyrLeuAlaPheSerTrpPheGluThrIleLeuLysAsnValGlyLysPr 138
    |||  :||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||
259 GGGAGTGCAGCAGCTTAATACAGCATCATTAATAATATACAAACGG 308
138 oLysValPheIleThrGlyPheAspValLysAsnGlyLysProAspPro. 154
    :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||
309 AAATGCTTACATGATTATATGTAATAATTAGTAAATAGATCAT 358
155 .....GlucIlyTrSerArgAlaArgAspLeuLeu 164
359 TACAATTAATGTTTGTAGAACAGCTTTACA...GCAAGTATGTTTTT 405
165 ArgGln.....AspLeuGlnLeuThrGlu 172
    |||||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||
406 AGACAAGCTACTAATACTATGGGAGATCTTCTCAGTTGCAGCATACAGG 455
172 yLysGlnAspLeuLysTyValValPhe...GluAspAlaProValGlyI 188
    :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||
456 AAGCGCTAGCTTACCTATATATTTGCTGAAGATGACCTACCACATG 505
188 IeLysAlaGlyLysAlaMetGlyAlaIleThrValGlyIleThrSerSer 204
    :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||
506 TAAAGACTGCTGACAAATTGATATAGTGTGCTTAACT...ATT 552
205 TyrAspLysSerValLeuPheAspAlaGlyAlaAspTyValValCysAs 221
    |||||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||
553 TATGATTTCACACACCTTGAGGAT..... 576
221 PleuThrGlnValSerValValLysAsnAsnGluAsn 233
    :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||
577 .....AAGATTGACAGAGCTTAGAACGATGCAAC 606

```

seq_name: /cgn2.6/prodata/2/lna/5B_COMB.seq:US-08-466-210A-14

seq_documentation_block:

Sequence 14, Application US/08466210A

Patent No. 5858362

GENERAL INFORMATION:

APPLICANT: Michel, James L.

APPLICANT: Kasper, Dennis L.

APPLICANT: Ausubel, Frederick M.

APPLICANT: Madoff, Lawrence C.

TITLE OF INVENTION: Conjugate Vaccine For Group B

TITLE OF INVENTION: Streptococcus

NUMBER OF SEQUENCES: 65

CORRESPONDENCE ADDRESS:

ADDRESS: Sterne, Kessler, Goldstein & Fox P.L.L.C.

STREET: 1100 New York Avenue, NW, Suite 600

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20005-3934

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/466,210A

FILING DATE: 06-JUN-1995

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/363,311

FILING DATE: 22-DEC-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/968,866

FILING DATE: 02-NOV-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/408,036

FILING DATE: 15-SEP-1989

ATTORNEY/AGENT INFORMATION:

NAME: Bugalsky, Lawrence B.

```

? REGISTRATION NUMBER: 35,086
? REFERENCE/DOCKET NUMBER: 0609.237000B
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (202) 371-2600
? TELEFAX: (202) 371-2540
? TELEX: 248636 SKK
? INFORMATION FOR SEQ ID NO: 14:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1380 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 79..1173
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: 1004
? OTHER INFORMATION: /note="this feature is to signify
? OTHER INFORMATION: that the nucleotide sequence from position 757
? OTHER INFORMATION: through 1003 is inserted at position 1004 and can be
? OTHER INFORMATION: repeated up to eight times (for a total of nine
? OTHER INFORMATION: repeating copies of these sequences within the
? OTHER INFORMATION: polynucleotide)."
US-08-466-210A-14

```

alignment_scores:

Quality:	109.50	Length:	229
Ratio:	0.928	Gaps:	11
Percent Similarity:	51.528	Percent Identity:	24.891

alignment_block:

US-09-558-284-2 x US-08-466-210A-14 ..

Align seq 1/1 to: US-08-466-210A-14 from: 1 to: 1380

```

27 AlaTrpThrLysLeuGlyTyrGlu...TyrGlyValAspProSerGluLe 42
    |||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||
31 CCTATATATTACTCTGTATAGACTTATACAGAGTAAAGGAAATATTAT 80
42 upheLysHisSer.....HisGlyAlaArgThrGlnGluValLeuArg 57
    :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||
81 GTTATAGAAGCTTAAATAATACACTTATGATTCACACAGCAACAAAC 130
57 rgpPhePheProLysLeuAspAspThrAspAsnLysGlyValLeuAlaLeu 73
    |||||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||
131 GCTTTCAATTAAAG..... 144
74 GluLysAspIleAlaHisSerTyrLeuAspThrValSerLeuIleProGlu 90
    |||||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||
145 .....AAGTTCAAGTTTGG 158
90 yAlaGluAsnLeuLeuSerLeuAsp...ValAspThrGlnThrGlnL 106
    |||||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||
159 TGCAGCTTCTGTACTAATGTGCTTAGTTTGGGGGGGTACACAAG 208
106 yLysLysLeuProGluArgLysTrpAlaIleValThrSerGlySer...Pro 121
    :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||
209 GTAATCTTAATATTATTGGAAGCAATAGTCTGCATCTACAAATGCA 258
122 TyrLeuAlaPheSerTrpPheGluThrIleLeuLysAsnValGlyLysPr 138
    |||  :||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||
259 GCGAGTGCAGCAGCTTAATACAGCATCATTAATAATATACAAACGG 308
138 oLysValPheIleThrGlyPheAspValLysAsnGlyLysProAspPro. 154
    :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||
309 AAATGCTTACATGATTATATGTAATAATTAGTAAATAGATCAT 358
155 .....GlucIlyTrSerArgAlaArgAspLeuLeu 164
359 TACAATTAATGTTTGTAGAACAGCTTTACA...GCAAGTATGTTTTT 405

```

165 Argln.....AspleuGlnLeuThrG1 172
|||||
406 AGACAAGTACTAAATACATGCGGATGTTTCAGTTCGACAGTACAGG 455
172 ylysglnaspleuLysTyrValValPhe..GluaspAlaProValGlyI 188
|||||
456 AAGGCTAGCTTACTACTATATATATTTGGTGAAGATGACCTACACATG 505
188 lelysalaclyLysAlaMetGlyAlaIleThrValGlyIleThrserser 204
|||||
506 TAAAGCTGATGACAAATTGATATAGTATGCTGCTTAACT...ATT 552
205 TyrAspLysSerValLeuPheaspAlaGlyAlaAspTyrValValCysas 221
|||||
553 TATGATTCAACAACCTTGAGGAT..... 576
221 pleuThrGlnValSerValValLysAsnAsnGluasn 233
|||||
577AAGATTGAAGAAGTTAGAACGAAATGCMAAC 606
seq_name: /cgn2_6/ptodata/2/lna/5B_COMB.seq:US-08-467-147A-14
seq_documentation_block:
Sequence 14, Application US/08467147A
Patent No. 5908629
GENERAL INFORMATION:
APPLICANT: Michel, James L.
APPLICANT: Kasper, Dennis L.
APPLICANT: Ausubel, Frederick M.
APPLICANT: Madoff, Lawrence C.
TITLE OF INVENTION: Conjugate Vaccine For Group B
TITLE OF INVENTION: Streptococcus
NUMBER OF SEQUENCES: 65
CORRESPONDENCE ADDRESSES:
ADDRESS: Sterne, Kessler, Goldstein & Fox P.L.L.C.
STREET: 1100 New York Avenue, NW, Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/467,147A
FILING DATE: 06-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/363,311
FILING DATE: 22-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/968,866
FILING DATE: 02-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/408,036
FILING DATE: 15-SEP-1989
ATTORNEY/AGENT INFORMATION:
NAME: Buglisky, Lawrence B.
REGISTRATION NUMBER: 35,086
REFERENCE/DOCKET NUMBER: 0609,2370009
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
TELEX: 248636 SKK
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 1380 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:

NAME/KEY: CDS
LOCATION: 79..1173
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1004
OTHER INFORMATION: /note="This feature is to signify
OTHER INFORMATION: that the nucleotide sequence from position 757
OTHER INFORMATION: through 1003 is inserted at position 1004 and can be
OTHER INFORMATION: repeated up to eight times (for a total of nine
OTHER INFORMATION: repeating copies of these sequences within the
US-08-467-147A-14
alignment_scores:
Quality: 109.50 Length: 229
Ratio: 0.928 Gaps: 11
Percent Similarity: 51.528 Percent Identity: 24.891
alignment_block:
US-09-558-284-2 x US-08-467-147A-14 ..
Align seg 1/1 to: US-08-467-147A-14 from: 1 to: 1380
27 ALATPThrLysLeuCySTyrGlu..TyrGlyValAlaSPProSerGlu 42
|||||
31 GCCTATAATTACTCTGTATAGAGTTATACAGAGTAAAGGAAATATTAT 80
42 uPheLysHisSer.....HisGlyAlaArgThrGlnGluValLeuArg 57
|||||
81 GTTTAGAGGCTTAAATAACAGTTATGATCTTACACAGCAGCAAC 130
57 rPhePheProLysLeuAspSPThrAspAsnLysGlyValLeuAlaLeu 73
|||||
131 GCTTTCAATTAA..... 144
74 GluLysAspIleAlaHisSerTyrLeuAspThrValSerLeuIlePro 90
|||||
145AAGTTCAAGTTTG 158
90 yAlaGluAsnLeuLeuSerLeuAsp..ValAspThrGluThrGln 106
|||||
159 TGCAGCTCTGTACTAATTGGTCTTATTTGGTGGGTTACACAG 208
106 yLysValPheProGluArgLysTrpAlaIleValThrSerGlySer...Pro 121
|||||
209 GTATCTTATATTTTGAAGAGTCAATGTTGCTGCATCTACAAATCCA 258
122 TyrLeuAlaPheSerTrpPheGluThrIleLeuLysAsnValGlyLysPr 138
|||||
259 GGCAGTGCAGCGACTTAATACAGCATCATCAAAATATACAAACCG 308
138 oLysValPheIleThrGlyPheAspValLysAsnGlyLysProAspPro 154
|||||
309 AATGCTTACATAGATTATATGATGTAATAATTGTAATAAAGATCCAT 358
155GluGlyTyrSerArgAlaIleArgAspLeuLeu 164
|||||
359 TACAATTAATTGTTTAGAACAAAGTTTACA...GCAAGATATGTTT 405
165 Argln.....AspleuGlnLeuThrG1 172
|||||
406 AGACAAGTACTAAATACATGCGGATGTTTCAGTTCGACAGTACAGG 455
172 ylysglnaspleuLysTyrValValPhe..GluaspAlaProValGlyI 188
|||||
456 AAGGCTAGCTTACTACTATATATTTGGTGAAGATGACCTACACAG 505
188 lelysalaclyLysAlaMetGlyAlaIleThrValGlyIleThrserser 204
|||||
506 TAAAGCTGATGACAAATTGATATAGTATGCTGCTTAACT...ATT 552
205 TyrAspLysSerValLeuPheaspAlaGlyAlaAspTyrValValCysas 221

|||||.....||| |||
553 TATGATTCACACACCTTGACGAT..... 576
221 pleuThrcInvalSerValValLyAsnAsnGluAsn 233
577AAGATTGAAGAGTTAGACGAATGCAAAAC 606
seq_name: /cgn2_6/prodata/2/lna/5B_COMB.seq:US-08-469-014-14

seq_documentation_block:
; Sequence 14, Application US/08469014
; Patent No. 5968521
; GENERAL INFORMATION:
; APPLICANT: Michel, James L.
; APPLICANT: Kasper, Dennis L.
; APPLICANT: Ausubel, Frederick M.
; APPLICANT: Madoff, Lawrence C.
; TITLE OF INVENTION: Conjugate Vaccine Against Group B
; NUMBER OF SEQUENCES: 65
; CORRESPONDENCE ADDRESS:
; ADDRESS: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, NW, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,014
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/363,311
; FILING DATE: 22-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/968,866
; FILING DATE: 02-NOV-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/408,036
; FILING DATE: 15-SEP-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Bugalsky, Lawrence B.
; REGISTRATION NUMBER: 35,086
; REFERENCE/DOCKET NUMBER: 0609.2370006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; TELEX: 248636 SSK
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1380 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 79..1173
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1004
; OTHER INFORMATION: /note="this feature is to signify
; OTHER INFORMATION: that the nucleotide sequence from position 757
; OTHER INFORMATION: through 1003 is inserted at position 1004 and can be
; OTHER INFORMATION: repeated up to eight times (for a total of nine
; OTHER INFORMATION: repeating copies of these sequences within the
; OTHER INFORMATION: polynucleotide)."
US-08-469-014-14

alignment_scores:
Quality: 109.50 Length: 229
Ratio: 0.928 Gaps: 11
Percent Similarity: 51.528 Percent Identity: 24.891

alignment_block:
US-09-558-284-2 x US-08-469-014-14 ..
Align seq 1/1 to: US-08-469-014-14 from: 1 to: 1380

27 AlatrphrlnLyLeuCystrGlu...TyrGlyValAspProSerGluLe 42
|||||.....||||| ||| |||
31 GCCTAATATTACCTCTGATAGAGTTATACAGAGTAAGAGCAATATTAT 80
42 uPhelYsnHsSer.....HisGlyAlaArgThrGlnGluValLeuAla 57
|||||..... |||..... |||
81 GTTTGAAGAGTCTAAATAATACAGTTATGATCTTCACAGACGAACAAC 130
57 rgrPheRheProLyLeuAspAspThrAspAsnLyGlyValLeuAlaLeu 73
||||| |||
131 GTTTTCATTTAG..... 144
74 GluLyAspRleAlaHisSerTyrLeuAspThrValSerLeuIleProG 90
145AAGTTCAAGTTTGG 158
90 yAlaGluAsnLeuLeuLeuSerLeuAsp...ValAspThrGlnThrGlnL 106
||||| |||..... |||
159 TGCACCTTCCTGACTAATTTGGCTTGTGGTGGGCTTACACAG 208
106 yLyLeuProGlnArgLyStrAlaIleValThrSerGlySer...Pro 121
||||| |||
209 GTAATCTTAATATTTTGAACAGTCATAGTTCGATCTTCAATTTCCA 258
122 TyrLeuAlaPheSerTyrPheGlnThrIleLeuLyAsnValGlyLysPr 138
||||| |||..... |||
259 GCGAGTCGACGACCTTAATATACAGCATCACTAAATAATATCAAAACGG 308
138 OLySValIheIleThrGlyPheAspValLyAsnGlyLysProAspPro. 154
||||| |||..... |||
309 AAATGCTTACATGATGATATGATGATGATGATGATGATGATGATGAT 358
155GluGlyTyrSerArgAlaArgAspLeuLeu 164
||||| |||..... |||
359 TACATTAATTTGTTTACACACAGTTTACA...GCCAAGATGTTT 405
165 ArgGln.....AspLeuGlnLeuThrG 172
||||| |||..... |||
406 AGACAAGGTAATAATTAATGAGGATGTTTCTCAGTTGCAGAGTACAGG 455
172 yLySGlnAspLeuLyTyrValValPhe...GluAspAlaProValGlyI 188
||||| |||..... |||
456 AAGGCTAGCTTCTTACCTTAATATATTTGGTGAAGATGAGTACACATG 505
188 IeLySAlaGlyLysAlaMetGlyAlaIleThrValGlyIleThrSerSer 204
||||| |||..... |||
506 TAAAGACTGATGACAAATTTGATATAGTTAGTTGCTTTAACT...ATT 552
205 TyrAspLySerValLeuPheAspAlaGlyAlaAspTyrValValCysAs 221
||||| |||..... |||
553 TATGATTCACACACCTTGACGAT..... 576
221 pleuThrcInvalSerValValLyAsnAsnGluAsn 233
577AAGATTGAAGAGTTAGACGAATGCAAAAC 606
seq_name: /cgn2_6/prodata/2/lna/PCTUS_COMB.seq:PCT-US93-10506A-14
seq_documentation_block:
; Sequence 14, Application PC/TUS9310506A
; GENERAL INFORMATION:
; APPLICANT: THE GENERAL HOSPITAL CORPORATION

APPLICANT: Fruit Street
APPLICANT: Boston, Massachusetts 02114
APPLICANT: United States of America
APPLICANT: 75 Francis Street
APPLICANT: Boston, Massachusetts 02115
APPLICANT: United States of America
TITLE OF INVENTION: Conjugate Vaccine Against Group B
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Ave., Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/10506A
FILING DATE: 02-NOV-1993
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/966,866
FILING DATE: 02-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: Cimbalia, Michele A.
REGISTRATION NUMBER: 33,851
REFERENCE/DOCKET NUMBER: 0609,237PC01
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
TELEX: 248636 SSK
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 1380 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 79..1173
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1004
OTHER INFORMATION: /note="This feature is to signify
that the nucleotide sequence from position 757
OTHER INFORMATION: through 1003 is inserted at position 1004 and can be
OTHER INFORMATION: repeated up to eight times (for a total of nine
OTHER INFORMATION: repeating copies of these sequences within the polynucleotide
PCT-US93-10506A-14
alignment_scores:
Quality: 109.50 Length: 229
Ratio: 0.928 Gaps: 11
Percent Similarity: 51.528 Percent Identity: 24.891
alignment_block:
US-09-558-284-2 x PCT-US93-10506A-14 ..
Align seg 1/1 to: PCT-US93-10506A-14 from: 1 to: 1380
27 AlATpThlyLeuCySTYrglu...TYrglyValAspProSerGluLeu 42
|||||: ||||| ||| ||| ||| : : :
31 GCCTAAATTTACTGTGTATAGAGTATACAGAGTAAGAGCAATATTAT 80
|||||: ||||| ||| ||| ||| : : :
42 ubpelyshIsSer.....HisGlyAlaArgThrgInGluValLeuArg 57
|||||: ||||| ||| ||| ||| : : :
81 GTTACAGAGTCTAAAAATACAGATTATGATCTTACACAGCAACAC 130
|||||: ||||| ||| ||| ||| : : :
27 AlATpThlyLeuCySTYrglu...TYrglyValAspProSerGluLeu 42
|||||: ||||| ||| ||| ||| : : :
31 GCCTAAATTTACTGTGTATAGAGTATACAGAGTAAGAGCAATATTAT 80
|||||: ||||| ||| ||| ||| : : :
42 ubpelyshIsSer.....HisGlyAlaArgThrgInGluValLeuArg 57
|||||: ||||| ||| ||| ||| : : :
81 GTTACAGAGTCTAAAAATACAGATTATGATCTTACACAGCAACAC 130
|||||: ||||| ||| ||| ||| : : :

57 rphpPheProLyLeuAspPThrAspAsnLySgLyValLeuAlaLeu 73
|||||: ||||| ||| ||| ||| : : :
131 GGTITTCATTAAG..... 144
74 GluLyAspPLeuAlaHisSerTyLeuAspThrValSerLeuIleProG 90
|||||: ||||| ||| ||| ||| : : :
145AAGTTCACAGTTTGG 158
90 yAlAGluAsnLeuLeuSerLeuAsp...ValAspThrGluThrgInL 106
|||||: ||||| ||| ||| ||| : : :
159 TGCACCTTCTGACTAATGCTTACTTTTGGTGCGGTTCACAG 208
|||||: ||||| ||| ||| ||| : : :
209 GTATCTTAATATTTTGAAGCATAGTGTGCAATCTACATTCACA 258
|||||: ||||| ||| ||| ||| : : :
106 ySlyLeuProGluArglySTypAlaIleValThrSerGlySer...Pro 121
|||||: ||||| ||| ||| ||| : : :
122 TyrLeuAlaPheSerTyrPheGluThrIleLeuLyAsnValGlyLysPr 138
|||||: ||||| ||| ||| ||| : : :
259 GGGAGTCGACGACCTTAATACAGCATCTAAATAATACAAACGG 308
|||||: ||||| ||| ||| ||| : : :
138 OlyValPheIleThrglyPheAspValLySasnGlyLysProAspPro 154
|||||: ||||| ||| ||| ||| : : :
309 AATGCTTACATGATTATATGATGATAAATAGCTAAATAGATCCAT 358
|||||: ||||| ||| ||| ||| : : :
155GluGlyTySerArgAlaArgAspLeuLeu 164
|||||: ||||| ||| ||| ||| : : :
359 TACATTTATTTGTTTGAACAGGTTTACA...GCAAGATGTTT 405
|||||: ||||| ||| ||| ||| : : :
165 ArgGln.....AspLeuGlnLeuThrgL 172
|||||: ||||| ||| ||| ||| : : :
406 AGACAAGGTACTAAATACTAGTGGAGTGTTCAGTTCCAGATCACAG 455
|||||: ||||| ||| ||| ||| : : :
172 yLySgInAspLeuLyTyValValPhe...GluAspAlaProValGlyI 188
|||||: ||||| ||| ||| ||| : : :
456 AAGGCTACTCTTACTATATATTTTGGTGAAGTGACATCACAGANG 505
|||||: ||||| ||| ||| ||| : : :
188 lElySalGlyLySalMetGlyAlaIleThrValGlyIleThrSerSer 204
|||||: ||||| ||| ||| ||| : : :
506 TAAAGACTGATGACAAATGTATATAGTGTGCTTAACT...ATT 552
|||||: ||||| ||| ||| ||| : : :
205 TyrAspLySerValLeuPheAspAlaGlyAlaAspTyValValCysAs 221
|||||: ||||| ||| ||| ||| : : :
553 TATGATTCAACAACCTTGAGGAT..... 576
221 pleuThrgInValSerValValLySasnGlnuAsn 233
|||||: ||||| ||| ||| ||| : : :
577AAGATTGAGAGATTAGAACGATGCAAC 606
|||||: ||||| ||| ||| ||| : : :
seq_name: /cgn2_6/ptodata/2/ina/PCTUS-COMB.seq:PCT-US93-10506-14
seq_documentation_block:
Sequence 14, Application PC/TUS9310506
GENERAL INFORMATION:
APPLICANT: Michel, James L.
APPLICANT: Kasper, Dennis
APPLICANT: Ausubel, Frederick M.
APPLICANT: Madoff, Lawrence C.
TITLE OF INVENTION: Conjugate Vaccine Against Group B
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Avenue, N.W., Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/10506
FILING DATE: 02-NOV-1993
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/968,866
FILING DATE: 02-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: Cimbalà, Michele A.
REGISTRATION NUMBER: 33,851
REFERENCE/DOCKET NUMBER: 0609.237PC01
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
TELEX: 248636 SSK
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 1380 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 79..1173
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1004
OTHER INFORMATION: /note="this feature is to signify
OTHER INFORMATION: that the nucleotide sequence from position 757
OTHER INFORMATION: through 1003 is inserted at position 1004 and can be
OTHER INFORMATION: repeated up to eight times (for a total of nine
OTHER INFORMATION: repeating copies of these sequences within the polynucleotide
PCT-US93-10506-14

alignment_scores:
Quality: 109.50 Length: 229
Ratio: 0.928 Gaps: 11
Percent Similarity: 51.528 Percent Identity: 24.891

alignment_block:
US-09-558-284-2 x PCT-US93-10506-14 ..
Align seg 1/1 to: PCT-US93-10506-14 from: 1 to: 1380

27 AlaTrpThrIysLeuGysTyrGlu...TyrGlyAlaAspProSerGluLe 42
|||||:|||||:|||||:|||||:
31 GCCATATATTACTCTGTATAGAGTTATACAGTAAGCAAGAAATATTAT 80
42 upbelyHisSer.....HisGlyAlaArgThrGlnGluValLeuArg 57
:|||||:|||||:|||||:|||||:
81 GTTAGAGAGGCTAAATAAATAACAGTTATGACTTCACACAGCAACAC 130
57 rghpethpPolysLeuAspAspThrAspAsnLysGlyValLeuAlaLeu 73
|||||:|||||:|||||:|||||:
131 GGTTCATTAATTAAG.....AAGTTCAAGCTTTGG 144
74 GluLysAspIleAlaHisSerTyrLeuAspThrValSerLeuIleProG 50
|||||:|||||:|||||:|||||:
145 .....AAGTTCAAGCTTTGG 158
90 yAlaGluAsnLeuLeuLeuSerLeuAsp...ValAspThrGluThrGln 106
|||||:|||||:|||||:|||||:
159 TGCAGCTCTGTACTAATGTGCTTACTTTTGGTGGGGTATACACAG 208
106 yslYsLeuProGluArgLysTrpAlaIleValThrSerGlySer...Pro 121
:|||||:|||||:|||||:|||||:
209 GTATATCTTAATATTTTTGAAGAGCAATAGTCTGCATCTACAAATTCA 258
122 TyrLeuAlaHisSerTrpPheGluThrIleLeuLysAsnValGlyLysP 138
|||||:|||||:|||||:|||||:
259 GCGAGTGCACGACCTTAATATACAGCATCAATAAATATACAAACCG 308
```

```

138 OlyValPheIleThrGlyPheAspValLysAsnGlyLysProAspPro. 154
:|||||:|||||:|||||:|||||:
309 AATGCTTACATAGATTATATGATGTAAATVTAGGTAAATAGATTCAT 358
155 .....GluGlyTyrSerArgAlaArgAspLeuLeu 164
:|||||:|||||:|||||:|||||:
359 TACATTAAATTGTTTATAGACAGAGTTTACA...GCCAAGTATGTTTTT 405
165 ArgGln.....AspLeuGlnLeuThrGln 172
|||||:|||||:|||||:|||||:
406 AGACAGGTACTAAATTAATGAGGATGTTCTCTCAGTTGACAGATCAGG 455
172 yLysGlnAspLeuLysTyrValValPhe...GluAspAlaProValGlyT 188
:|||||:|||||:|||||:|||||:
456 AAGGCTAGTCTTACCTATATATATTTGGTGAAGATGAGCTACACATG 505
188 IeLysAlaGlyLysAlaMetGlyAlaIleThrValGlyIleThrSerSer 204
:|||||:|||||:|||||:|||||:
506 TAAAGACTGATGACCAAAATGATATAGTTAGTGTGCTTAACT...ATT 552
205 TyrAspLysSerValLeuPheAspAlaGlyAlaAspTyrValValCysAs 221
|||||:|||||:|||||:|||||:
553 TATGATTCAACACCTTGAGGAT..... 576
221 pleuThrGlnValSerValValLysAsnAsnGluAsn 233
:|||||:|||||:|||||:|||||:
577 .....AAGATTGAAGCAAGTTAGACGATGCAAC 606

seq.name: /cgn2_6/prodata/2/ina/5h_comb.seq:US-08-937-102-1
seq.documentat_block:
: Sequence 1, Application US/08937102A
: Patent No. 5965134
: GENERAL INFORMATION:
: APPLICANT: Thiel, Heinz-Jurgen
: APPLICANT: Elpers, Knut
: TITLE OF INVENTION: T Cell Stimulating Protein of Pestivirus
: FILE REFERENCE: I/94108
: CURRENT APPLICATION NUMBER: US/08/937,102A
: EARLIER FILING DATE: 1997-09-24
: EARLIER APPLICATION NUMBER: 08/693,247
: NUMBER OF SEQ ID NOS: 36
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 3639
: TYPE: DNA
: ORGANISM: Classical Swine Fever Virus
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (1)..(3639)
US-08-937-102-1

alignment_scores:
Quality: 91.00 Length: 300
Ratio: 0.599 Gaps: 21
Percent Similarity: 50.667 Percent Identity: 24.333

alignment_block:
US-09-558-284-2 x US-08-937-102-1 ..
Align seg 1/1 to: US-08-937-102-1 from: 1 to: 3639

7 AspleuGysLeuPheAspLeuAspGlyThrIleValSerThrThrValAl 23
:|||||:|||||:|||||:|||||:
2253 AACCTATGCCCTTACGGTTA.....CTTTGCCAGATGTACAAACCA 2296
23 AlaGluLys.....AlaTrpThrL 30
|||||:|||||:|||||:|||||:
2297 AGCTGAGCGCGCAATGGTAGAATATTCCTTATATATTCCTAGATGAGTAT 2346
```



```

KEYWORDS      AZ683512.1  GI:11820658
SOURCE        Entamoeba histolytica.
ORGANISM      Entamoeba histolytica.
REFERENCE     1 (bases 1 to 921)
AUTHORS       Loftus,B., Van Aken,S. and Fraser,C.
TITLE         Determination of clone end sequences from Entamoeba histolytica
              HMI:IMSS sheared DNA library
JOURNAL       Unpublished (2000)
COMMENT       Contact: Brendan J Loftus
              Department of Eukaryotic Genomics
              The Institute for Genomic Research
              9712 Medical Center Dr., Rockville, MD 20850, USA
              Tel: 301 838 0200
              Fax: 301 838 0208
              Email: bjoftus@stgfr.org
              Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
              DNA library
              Seq primer: M13-Forward
              Class: shotgun
              High quality sequence start: 29
              High quality sequence stop: 769.
              Location/Qualifiers
                1..921
                  /organism="Entamoeba histolytica"
                  /strain="HMI:IMSS"
                  /db_xref="taxon:5759"
                  /clone_lib="Entamoeba histolytica Sheared DNA"
                  /note="Vector: PHOS1. Site 1: Bst I; Constructed at The
                  Institute for Genomic Research (TIGR), Rockville, MD.
                  Genomic DNA isolated from broth cultures of E. histolytica
                  using a method described by Clark and Diamond (Clark,
                  C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
                  method for isolate identification. Exp. Parasitol.
                  77:450.). The DNA was mechanically sheared to give a
                  tight size distribution (~2 kb). The v + i method used for
                  the library construction is described in detail in Smith
                  H.O. and Venter, J.C. (Making small insert libraries for
                  whole genome shotgun sequencing projects. In Genome
                  Sequencing: A Practical Approach, eds. M. Vaudin and B.
                  Barrell, Oxford University Press, 1999)."
```

461 TATTAAATGACACACTTGACATCATACATATAATAATGATCATTTAA 412
80 ..SerTYrLeuAspThrValSerLeuIleProGlyIaGluAsnLeu 95
411 ATAAATTATGTCACACAGTAAACCATTTACAGAGACCATGAAGATTTA 362
96LeuSerLeuAspValAspThrGluThrG1 105
361 AATTATTTTAAAAACATCATCTTCATGCTATGCTGTACTTCAACGCAC 312
105 nLysIysLeuProGluArgLysTrp..... 113
311 TAAATCTGCTTTGACAAAAAATGTTAAAAACCAAGATGCTCATTT 262
114AlaIleValThrSerGlySerProTYrLeuAlaIleSerTrp 127
261 APTTTGATGCAAATTGTTCTTGAGATGATCCTCAT..... 227
128 pheGluThrIleLeuLysAsnValGlyLysProLysValPheIleThrG1 144
227 227
144 yPheAspValLysasnGlyLysProAspProGluGlyTYrSerArgAlaA 161
226GTAAAGACGCTAAACCAATCCACAAATATTTTTCATGCTG 184
161 rgAspLeuLeuArgGlnAspLeuGlnLeuThrGlySglnAspLeuLys 177
183 GTCATTTACTT.....GGATGCTAGTCACATGAA 155
178 Tyr...ValValPheGluAspAlaProValGlyIleLysAlaGlyLysAl 193
154 CACAGCAATGTTGTAAGATGATGCTACTTGCTGCTTCAACACAGAAATGC 105
193 aMetGlyAlaIleThrValGlyIle 201
104 ATCTGACCTTATCTACTGACTTAT 80
seq_name: gb_est17:AI209756
seq_documentation_block:
LOCUS AI209756 266 bp mRNA EST 19-Oct-1998
DEFINITION C9b02a1.r1 Aspergillus nidulans 24hr asexual developmental and
vegetative cDNA lambda zap library Aspergillus nidulans cDNA clone
c9b02a1 5', mRNA sequence.
ACCESSION AI209756
VERSION AI209756.1 GI:3771698
KEYWORDS EST.
SOURCE Aspergillus nidulans.
ORGANISM Aspergillus nidulans
REFERENCE Eukaryota: Fungi; Ascomycota: Pezizomycotina; Eurotiomycetes;
Aurotiales: Trichocomaceae; Emeritella.
1 (bases 1 to 266)
AUTHORS Knaple,D., Gray,J., Hausner,J., Lai,H., Martin,W., Aramayo,R.,
Prade,R. and Roe,B.
TITLE An Aspergillus nidulans EST Database
COMMENT Unpublished (1998)
Other ESTs: c9b02a1.f1
Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu
Department of Chemistry and Biochemistry
Advanced Center for Genome Technology, University of Oklahoma
620 Parrington Oval, Norman, OK 73019, USA
Tel: 405 325 4912
Fax: 405 325 7762
Email: broe@ou.edu
We anticipate the future release of the cDNA clones to the Fungal
Genetics Stock Center
Seq primer: T3
High quality sequence stop: 161.
location/Qualifiers
1..266
/organism="Aspergillus nidulans"


```
/strain="FGSC A26"  
/db_xref="taxon:5072"  
/clone_lib="c9b02a1"  
/clone_lib="Aspergillus nidulans 24hr asexual  
developmental and vegetative cDNA lambda zap library"  
/tissue_type="vegetative mycelia, asexual structures"  
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:  
XhoI; 5' end of cDNA cloned into EcoRI site of pBluescript  
3' end of cDNA cloned into XhoI site of pBluescript"  
BASE COUNT      56 a      66 c      84 g      60 t  
ORIGIN  
  
alignment_scores:  
  quality: 139.00      length: 97  
  ratio: 2.242      gaps: 3  
  percent similarity: 63.918      percent identity: 35.052  
  
alignment_block:  
us-09-558-284-2 x AI209756 ..  
  
Align seg 1/1 to: AI209756 from: 1 to: 266  
  
113 TTPAlaIleValIThrSerGlySerProTyrIleuAlaPheSerTrpPheG1 129  
1 TGGGGCGCTTGACGTCGGGACCGCGCTCATTTGACGGGTGGTGG 50  
129 uThrIleLeuLysAsnValGlyLysProLysValPheIleThrGlyPhea 146  
51 TGTTCCTC.....AAGCTCACACATCCAGATGCTCTCTTGGCGAGG 94  
146 sPValLysAsnGlyLysProAspProGluGlyTyrSerArgAlaArgAsp 162  
95 ATGTGAAGACTCGGAGACGAGATCCGGATGCTTACCTGTGACGAAG 144  
163 LeuLeuArgGlnAspLeuGlnLeuThrGlyLysGlnAspLeuLysTyr. 178  
145 .....AAATGGGTCTAGAACATTC 164  
179 .....ValValPheGlnAspAlaProValGlyIleLysAlaGlyL 192  
165 CTCCTCATACTAGTGGTGGAGGAGACGCCCTCAGGCAATAAGGCTGCA 214  
192 ysaLametGlyAlaIleThrValGlyIleThrSerSerTyr 205  
215 AGCGACGCCGTTTACATTCATTCCTTTGACACCACTCAG 255  
  
seq_name: gb_estc97:BG129500  
  
seq_documentation_block:  
LOCUS      BG129500      717 bp      mRNA      EST      31-JAN-2001  
DEFINITION EST475146 tomato shoot/meristem Lycopersicon esculentum cDNA clone  
COTF24P11 5' sequence. mRNA sequence.  
ACCESSION      BG129500  
VERSION      BG129500.1 GI:12629688  
KEYWORDS      EST.  
SOURCE      tomato.  
ORGANISM      Lycopersicon esculentum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asterales; eunasterids I; Solanales; Solanaceae; Solanum;  
Lycopersicon.  
1 (bases 1 to 717)  
van der Hoeven,R., Bezzerides,J., Sun,H., Cho,J., Ulterback,T.,  
Hansen,C., Ronning,C. and Tanksley,S.  
Generation of ESTs from tomato shoot/meristem tissue  
Unpublished (2001)  
Contact: CUGI  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Email: http://www.genome.clemson.edu/orders/index.html.  
Location/Qualifiers
```

```
source      1. 717  
/organism="Lycopersicon esculentum"  
/cultivar="TA496"  
/db_xref="taxon:4081"  
/clone="cOTF24P11"  
/clone_lib="tomato shoot/meristem"  
/tissue_type="shoot/meristem"  
/dev_stage="developing shoots from 4-6wks old plants"  
/lab_host="SOLR"  
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:  
XhoI; Small expanding leaves from the growing tip were  
taken from greenhouse plants (4-6wks old TA496). Tissue  
was immediately frozen in liquid nitrogen."  
BASE COUNT      208 a      129 c      177 g      203 t  
ORIGIN  
  
alignment_scores:  
  quality: 139.00      length: 205  
  ratio: 1.311      gaps: 7  
  percent similarity: 51.707      percent identity: 27.317  
  
alignment_block:  
us-09-558-284-2 x BG129500 ..  
  
Align seg 1/1 to: BG129500 from: 1 to: 717  
  
10 LeuPheAspLeuAspGlyThrIleValSerThrValAlaIleGlyL 26  
69 CTATTTGATGATGAAGGAGATGATGATGATGATGATGATGATGATGAT 118  
26 sAlaTrpThrLysLeuGlyTyrGlyTyrGly.....V 37  
119 TGCCTTCGCGTGAATGCTTTCAGAGATAGCGTACATCAGGCTGCCAG 168  
37 aLAspProSerGlyLeuPheLysHisSerHisGlyAlaArgThrGlnGlu 53  
169 TTGATGAGAGAGTGTTCATCAAGACTATTGCTGGGAAACACACAGATAT 218  
54 ValLeuArgArgPhePheProLysLeuAspPheArgAspAsnLysGlyVa 70  
219 ATTGCTTTCGCCCTTTTTCCT.....GATGATCAGGAACGGGCTTTAA 262  
70 lLeuAlaLeuGlnLysAspIleAlaHisSerTyrLeuAspThrValSerL 87  
263 ATTTTGTATGATGAAGGAGATGATGATGATGATGATGATGATGATGAT 312  
87 euIleProGlyAlaGlnAsnLeuLeuSerLeuAspValAspThrGlu 103  
313 TGAACCT.....ATGATGGTTTGTACAG 338  
104 ThrGlnLysLysLeuProGluArg.....LysTrpAlaIleValThrSe 118  
339 GTGAAGAAATGATTTGAAGATGCTGACCTGAACCGCTGACATTACAA 388  
118 rGlySerProTyrLeuAlaPheSerTrpPheGluThrIleLeuLysAsnV 135  
389 TGCACCTAGACTAAAGCT.....GAACCTGATATAGAAATAC 426  
135 aIGlyLysProLysValPhe.....IleThrGlyPheAspValLys 148  
427 TTGGCCTAAAGGATTTCTTGATGATGATGATGATGATGATGATGATG 476  
149 AsnGlyLysProAspProGluGlyTyrSerArgAlaArgAspLeuArg 165  
477 CGTGCAAAACCATCTCTGACCTTACTTTGAAGCGCTTGACAGCTTAA 526  
165 gGlnAspLeuGlnLeuThrGlyLysGlnAspLeuLysTyrValValPheG 182  
527 GGTGTCTAAGGACACAG.....TTCAATATTTG 555  
182 lAspAlaProValGlyIleLysAlaGlyLysAlaMetGlyAlaIleThr 198
```

149 TTTGTTTGAATTTTAAATGGACCTTAAATATTTGATATCTCTCTCAATCCATTTT

Prade, R. and Roe, B.

TITLE An Aspergillus nidulans EST Database
JOURNAL Unpublished (1998)
COMMENT Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu
Department of Chemistry and Biochemistry
Advanced Center for Genome Technology, University of Oklahoma
620 Parrington Oval, Norman, OK 73019, USA
Tel: 405 325 4912
Fax: 405 325 7762
Email: broe@ou.edu
We anticipate the future release of the cDNA clones to the Fungal
Genetics Stock Center
Seq primer: SK
High quality sequence stop: 237.
Location/Qualifiers
1. 263
/organism="Aspergillus nidulans"
/strain="FGSC A26"
/db_xref="taxon:5072"
/clone="A1B02C9"
/clone_11b="Aspergillus nidulans 24hr asexual
developmental and vegetative cDNA lambda zap library"
/tissue.type="vegetative mycelia, asexual structures"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; 5' end of cDNA cloned into EcoRI site of pBluescript
3' end of cDNA cloned into XhoI site of pBluescript"

BASE COUNT 56 a 67 c 80 g 60 t
ORIGIN

alignment_scores:
Quality: 129.00 Length: 83
Ratio: 2.481 Gaps: 2
Percent Similarity: 62.651 Percent Identity: 39.759

alignment_block:
US-09-558-284-2 x AA788532 ..

Align seq 1/1 to: AA788532 from: 1 to: 263

```

138 ProlysaValPheIleThrGlyPheAspValIysAsnGlyLysProAspPr 154
|||||.....:|||||:|||||
43 CCACATGCTCCTGCTGTCGCGAGCATGTAGACTCCGGAAGCCAGATCC 92
154 oGUGLyTYrSerArgAlaArgAspLeuArgGlnAspLeuGlnLeuT 171
|||||.....:|||||
93 GCGATGCTACTGCTGTCGACCAAG..... 117
171 hrgLyLysGlnAspLeuLysTYr.....ValValPheGlnAsp 183
|||||.....:|||||
118 ....AAATGGGTCTAGAACATTCTCTCGATAGTGTCTGTCGAGAGAT 162
184 A1AProValGlyIleLysAlaGlyLysAlaMetGlyAlaIleThrValG 200
|||||.....:|||||
163 GCCGCCGTCAGCATAAAGCTGCGACAGCCGCTTCACATGTCATTGC 212
200 yIleThrSerSerTYrAspLysSerValLeuPheAspAlaGlyAlaAsp 216
|||||.....:|||||
213 TTTCACCACTCACACTTTGAGACAGCTTCAACGCCGTCGCGCGAGC 261
seq_name: gb_est51:AW774770
seq_documentation_block:
LOCUS AW774770 729 bp mRNA EST 07-SEP-2000
DEFINITION EST333921 KV3 Medicago truncatula cDNA clone pkv3-23L2. mRNA
sequence.
ACCESSION AW774770
VERSION AW774770.1 GI:7718687
KEYWORDS EST.
SOURCE barrel medic.
ORGANISM Medicago truncatula
Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots:
Rosidae: eurosids I: Fabales: Fabaceae: Papilionoideae: Medicago.

```

REFERENCE 1 (bases 1 to 729)
AUTHORS VandenBosch,K., Hurt,J., Moore,J., Beremand,P., Peng,H., Ellis,L.,
Town,C.D., Bowman,C.L., Craven,M.B., Hansen,T.S., Holt,I.E. and
Fraser,C.M.
JOURNAL Unpublished (1999)
COMMENT ESTs from roots of Medicago truncatula after Rhizobium inoculation
Contact: VandenBosch K
Department of Biology
Texas A&M University
College Station, TX 77843-3258, USA
Tel: 409 845 7707
Fax: 409 845 2891
Email: kate@mail.bio.tamu.edu
Texas A&M EST name: T258189e
TIGR sequence name: MREBF617K
More information is available at:
http://chrtsie.tamu.edu/medicago
Seq primer: Skmod (CTA gaa cta gta gta gta CC).
Location/Qualifiers
1. 729
/organism="Medicago truncatula"
/cultivar="genotype A17"
/db_xref="taxon:3880"
/clone="pkv3-23L2"
/clone_11b="KV3"
/tissue.type="Seedling roots"
/dev.stage="3 days post inoculation with Sinorhizobium
meliloti"
/lab_host="E. coli strain XL0LR"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA
was directionally ligated into the XhoI/XhoI vector from
Stratagene and packaged using GigaPack III Gold packaging
extracts. Plasmids containing cDNA inserts were excised
from the recombinant lambda-zap phage using Ex-assist
helper phage and propagated in XL0LR cells."

BASE COUNT 209 a 133 c 167 g 220 t
ORIGIN

alignment_scores:
Quality: 127.00 Length: 223
Ratio: 1.085 Gaps: 9
Percent Similarity: 52.466 Percent Identity: 26.009

alignment_block:
US-09-558-284-2 x AW774770 ..

Align seq 1/1 to: AW774770 from: 1 to: 729

```

10 LeuPheAspLeuAspGlyThrIleValSerThrThrValAlaIleGly 26
|||||.....:|||||
46 TTGTTGATATAGACGGAAGCTTTGTGATTCGACCAATCCACTATA 95
26 sAlaTrpThrLysLeuGlyTYrGlyValAsp..... 38
|||||.....:|||||
96 TGCCTTTCGTAATGCTTCAAGAGATGCTTCAATGAGAGCAATCCTA 145
39 ..ProSerGlnLeuPheLysHisSer...HisGlyAlaArgThrGlnGlu 53
|||||.....:|||||
146 TTAACAGAGAAATTTTATTCGCACTGTCGTGGCAACCAAGATGATAC 195
54 ValLeuArgArgPhePheProLysLeuAspPheTrpAspLysGlyVal 70
|||||.....:|||||
196 ATTGCTTTGATCTCTTCC.....GGTATCGGAGCGAGGTTTAA 239
70 LLeuAlaLeuGlnLysAspIleAlaHisSerTYrLeuAspThrValSerL 87
|||||.....:|||||
240 GTTGTAGAAAGATAAAGACCATGTTCCGAGATTTAGCACCTGACGAC 289
87 euLleProGlyAlaGlnAsnLeuLeuSerLeuAspValAspThrGlu 103
|||||.....:|||||
290 TGAAGCCG.....TTGAATGGTCTTGAT.....AAA 315

```



```

142 GAGTAACCAACTGATTTGAAGATGCATTAAATGAGTACGAAGCAGCAT 93
192 ysal1amelc1yala1etnva1g1l1elhserferiyasplys 207
||||| |||||:::||:::||:::||:::||:::||:::||:::||:::||:::||
92 TGGCACAGCGCATTCAGCAATTCCTATACGACATTCATATTAAA 46

seq_name : gb_gss5.CNS06DY4

seq_documentation_block:
LOCUS      CNS06DY4          808 bp    DNA             GSS              05-APR-2001
DEFINITION T3 end of clone AR0A017A03 of library AR0A from strain CBS 732 OT
ACCESSION  AL394418
VERSION    AL394418.1 GI:12145375
KEYWORDS   GSS: clone AR0A017A03; RST AR0A017A03TP1.
SOURCE     Zygosaccharomyces rouxii.
ORGANISM   Zygosaccharomyces rouxii.
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Basidiomycetales; Saccharomycetaceae; Zygosaccharomyces.
REFERENCE  1 (bases 1 to 808)
AUTHORS   de Montigny,J., Straub,M., Potier,S., Tekata,F., Dujon,B.,
            Wincker,P., Artiguenave,F. and Souciet,J.
TITLE     Genomic exploration of the hemiascomycetous yeasts: 8.
           Zygosaccharomyces rouxii
JOURNAL   FEBS Lett. 487 (1), 52-55 (2000)
MEDLINE   20584718
REFERENCE  2 (bases 1 to 808)
AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Boitard-Pukuhara,M., Bon,E., Brotier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Jeplinge,A., Llorente,B.,
            Maitelly,A., Neuveglise,C., Ozler-Kalogeropoulos,O., Potier,S.,
            Sarrin,P., Tekata,F., Toftano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
TITLE     Genomic exploration of the hemiascomycetous yeasts: 1. A set of
           yeast species for molecular evolution studies
JOURNAL   FEBS Lett. 487 (1), 3-12 (2000)
MEDLINE   20584711
REFERENCE  3 (bases 1 to 808)
AUTHORS   Genoscope.
TITLE     Direct Submission
JOURNAL   Submitted (28-Mar-2001) Genoscope - Centre National de Sequencage,
           2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
           seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENTARY This BAC end sequence is part of a random genomic sequencing
           program of thirteen yeast species:
           . Saccharomyces bayanus var. uvarum,
           . Saccharomyces exiguus,
           . Saccharomyces servazii,
           . Zygosaccharomyces rouxii,
           . Saccharomyces kluyveri,
           . Kluyveromyces thermotolerans,
           . Kluyveromyces lactis var. lactis,
           . Kluyveromyces marxianus var. marxianus
           . Pichia angusta,
           . Debaryomyces hansenii var. hansenii,
           . Pichia sorbitophila,
           . Candida tropicalis and
           . Yarrowia lipolytica.
           Genomic inserts of 3 to 5 kb were prepared and both extremities
           were sequenced. See keywords for description of this sequence and
           for the sequence of the other extremity of this insert.
FEATURES
source
1..808
/organism="Zygosaccharomyces rouxii"
/strain="CBS 732"
/db_xref="taxon:4956"
/cclone="AR0A017A03"
/cclone_1id="AR0AA"
/note="end : T3"
/complement(<3..>370)
/note="similar to Saccharomyces cerevisiae ORF YIL053w [
RHR2 ; DL-glycerol phosphatase ]
2 putative frameshift(s)"

```

```

misc_feature
complement(<3..>358)
/note="Similar to Saccharomyces cerevisiae ORF YER062c [
HOR2 ; DL-glycerol phosphate ]
2 putative frimessilit(s)"
/evidence=not_experimental

BASE COUNT      245 a      160 c      188 g      210 t      5 others
ORIGIN

alignment_scores:
      Quality: 123.00      Length: 119
      Ratio: 1.757      Gaps: 6
Percent Similarity: 58.824      Percent Identity: 33.613

alignment_block:
US-09-558-284-2 x CNS06DY4/rev ..

Align seg 1/1 to reverse of: CNS06DY4 from: 1 to: 808

9  CysLeuPheAspLeuPheAspGlyThrIleVal.SerThrThrValAlaIaIaG 25
|||||.....:|||||.....:|||||
316 TGCTCTTTGACGCTTGACGCTACATCATCATCTCCCAANCCAGCTATGCG 267
25  IuIaSaIaTtPThrIuIaCysTyrgIuTyrgIy.....:..... 36
|||
266 CTGATCTTCTGAGA.....GAATTCGCTAAGGACAAAGCCTTAC 229
37  ValAspProSerGluLeuPheIuHisSerHisGlyAlaIaArgThrIuIaG 53
||| : : : : : : : : : : : : : : : : : : : : : : : : :
228 TTGCACTTCCCAATACGTCATCAACATCTCTACAGTTGGAGACTTAAAGN 179
53  uValLeuArg.ArgPhePheProIuIaAspAspPThrAspAsnIuIaGly 69
|||||.....:|||||.....:|||||
178 AGCCTTTCGCTAAGTTTCGCTCCA.....GACTACGCTGATGAAGAATAC 135
70  ValLeuAlaLeuGluIuIaAspIleAlaHisSerTyrrLeuAspThrValSe 86
: : : : : : : : : : : : : : : : : : : : : : : : :
134 GCTACCAAGGTGAGAGCTGAATCCGACACAAATATAGGTCAAGTCCAT 85
86  rLeuIleProGlyAlaGluAsnLeuLeuSerLeuAspValAspThrG 103
: : : : : : : : : : : : : : : : : : : : : : : : :
84 CGAAGTCGCCGGTGTACAGAGCTTGCTAAGCTTTGAACGCT..... 42
103 IuThrIuIaIuIaCysLeuProGluIaArgLysTrpAlaIleValThrSerGly 119
|||||.....:|||||.....:|||||
41  .....TTGCCAAAGGAAATAATGGCGCTTGCTACTTCCGCT 6
120 Ser 120
|||
5 TTC 3

seq_name: gb_est102:BG589791

seq_documentation_block:
LOCUS      BG589791      728 bp      mRNA      EST      12-APR-2001
DEFINITION ES1497633 P. infestans-challenged leaf Solanum tuberosum cDNA clone
BPLI3C1 5' sequence, mRNA sequence.
ACCESSION  BG589791
VERSION    BG589791.1 GI:13607931
KEYWORDS   EST.
SOURCE     potato.
ORGANISM   Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliopsida; eudicotyledons; core eudicots;
Asteridae; easterids I; Solanales; Solanaceae; Solanum.
1 (bases 1 to 728)
Zhang,P., Hernandez,M., Torunqvist,C.-E, Wirtz,U., Loukianov,A.,
Rangel,P., Haberland,G.T., Cho,J., Chilingiro,A., Bougri,O., Bueli
,C.R., Renning C.M., Helgeson,J. and Baker,B.
Generation of ESTs from Potato leaves Challenged with Phytophthora
Infestans. Incompatible Reaction
Unpublished (2000)
JOURNAL

```

COMMENT

Contact: Cathy Rinning
The Institute for Genomic Research
For clone info: Please contact Research Genetics, Libraries
Division tel 1-800-711-6195, email cdna@resgen.com
Seq primer: M13F-R.
Location/Qualifiers

FEATURES
source

1. 728
/organism="Solanum tuberosum"
/cultivar="Kennebec"
/db_xref="taxon:4113"
/clone="BPL1301"
/issue_id="P. Infestans-challenged leaf"
/issue_type="leaf"
/dev_stage="6 week old"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-). Site_1: EcoRI; Site_2:
XhoI. Whole plants were challenged with 450,000
sporangia/ml P. infestans US-1(US 940501) in Biotron
(Madison, Wisconsin). Leaf tissue was collected at 1, 2,
5, 12, and 24 hours post-challenge and frozen in liquid
nitrogen immediately upon removal. Kennebec plants showed
no signs of HR. Kathdin plants (susceptible to P.
infestans US-1) were used as controls and showed
infection. NOTE: We cannot exclude the possibility that
this sequence is actually derived from Phytophthora rather
than potato."

BASE COUNT 190 a 148 c 177 g 213 t
ORIGIN

alignment_scores:

Quality: 121.00 Length: 203
Ratio: 1.210 Gaps: 7
Percent Similarity: 49.261 Percent Identity: 24.138

alignment_block:

US-09-558-284-2 x BG589791 ..

Align seg 1/1 to: BG589791 from: 1 to: 728

```

34 GltutrgllyValAspProSerGluLeuPheLysHisSerHisGlyAlaAr 50
   ::::::::::::::::::::
33 GATTATGGCGCTCCACATCTCA..... 53
50 gThGluGluValLeuArgArgPhePheProLysLeuAspAspThrAspa 67
   ::::::::::::::::::::
54 .ACGATGCAATTCTTCGCAATTTTACCAATGCTGCTGAC..... 95
67 snLysGlyValLeuAlaLeuGluLysAspIleAlaHisSerTyrLeuAsp 83
96 ..... 104
84 ThrValSerLeuLeuProGlyAlaGluAsnLeu..... 95
105 AATATCAAGCTGTCCAGGGCCAAATCGATTGATTACCATTTGAGGG 154
96 .....LeuSerLeuAspValAspThrGluThrGlnLysLysLeuProG 110
155 TCATGGGTACCATGGCATGGCGTCAATTTCTTAAGTCAACATAG 204
110 LuanGlyStrpAlaIleValThrSerGlySerProTyrLeuAlaPheSer 126
205 AGGCCCAA.....ATTTTCATCATGCAGCA 230
127 TrpPheGluThrIleLeuLysAsnValGlyLysProLysValPheIleTh 143
231 TCGAAGAAATCCCTTCACGCTATGTTGA..... 260
143 rglYpheaSpValLysAsnGlyLysProAspProGluGlyTyrSerArg 160
261 .GGAGCGAGTGAAGAGTGAAGAGCCATCTCCGAAATATTTCTTGAAG 309
160 laArgAspLeuLeuArgGlnAspLeuGlnLeuThrGlyLysGlnAspLeu 176

```

```

310 CAGCTAAAGACATCAACATGAT.....CCATCC 338
177 LysTyrValValPheGluAspAlaProValGlyIleLysAlaGlyLysAl 193
   ::::::::::::::::::::
339 AGCTTTCTTCATTTGAGATTCATACACAGAGTGTACTCTGCTAAGGC 388
193 aMetGlyAlaIleThrValGlyIleThrSerSerTyrAspLysSerVal 210
389 TGCTGGAATGCGACAGTGTGCTGACCATCTTCGCAACAGATGATC 438
210 eupheaSpAlaGlyAlaAspTyrValValLysAspLeuThrGlnValSer 226
439 TTTAT...ACTTCGTGATGATGATCAATTCCTCTAGATTGCAA 485
227 ValValLys 229
486 CTAGAAAG 494

```

seq_name: gb_gss30:A2682029

seq_documentation_block:

LOCUS A2682029 852 bp DNA GSS 14-DEC-2000
DEFINITION ENTHP38TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
genomic DNA sequence.

ACCESSION A2682029 GI:11819175

VERSION A2682029.1 GI:11819175

KEYWORDS GSS.

SOURCE Entamoeba histolytica.

ORGANISM Entamoeba histolytica.

REFERENCE 1 (bases 1 to 852)

AUTHORS Loftus/B., Van Aken,S. and Fraser,C.

TITLE Determination of clone end sequences from Entamoeba histolytica

HML:IMSS sheared DNA library

UNPUBLISHED (2000)

CONTACT: Brendan J Loftus

DEPARTMENT: The Institute for Genomic Research

7972 Medical Center Dr., Rockville, MD 20850, USA

TEL: 301 838 0200

FAX: 301 838 0208

EMAIL: bjloftus@tigr.org

CLONES: Clones are derived from the Entamoeba histolytica HML:IMSS sheared

DNA library

Seq primer: M13-Reverse

CLASS: Shotgun

High quality sequence start: 15

High quality sequence stop: 836.

FEATURES
source

1. 852

/organism="Entamoeba histolytica"

/strain="HML:IMSS"

/db_xref="taxon:5759"

/clone_lib="Entamoeba histolytica Sheared DNA"

/note="Vector: pHOSt; Site_1: Bat I; Constructed at The

Institute for Genomic Research (TIGR), Rockville, MD.

Genomic DNA isolated from broth cultures of E. histolytica

using a method described by Clark and Diamond (Clark,

C.G., and Diamond, L.S., (1993) Entamoeba histolytica: a

method for isolate identification. Exp. Parasitol.

77:450.). The DNA was mechanically sheared to give a

tight size distribution (~2 kb). The v + i method used for

the library construction is described in detail in Smith,

H.O. and Venter, J.C. (Making small insert libraries for

whole genome shotgun sequencing projects. In Genome

Sequencing: A Practical Approach, eds. M. Vaudin and B.

Barell, Oxford University Press, 1999).

BASE COUNT 287 a 120 c 114 g 331 t
ORIGIN

alignment_scores:

Quality: 121.00 Length: 207
Ratio: 1.198 Gaps: 9
Percent Similarity: 48.792 Percent Identity: 24.638

alignment block:
US-09-558-284-2 x AZ682029/rev ..

Align seg 1/1 to reverse of: AZ682029 from: 1 to: 852

```

28 TrpThrIleuSerGlyGluTyrGlyValAspProSerGluLeuPheLeu 44
   ||||| ..... |||||
851 TATACAAATTGTTACTCAACCAATAT.....TTACATGA 820
   ||||| ..... |||||
44 ShisSerHisGlyAlaArg...ThrGlnGluValLeuArgArgPhePhe 60
   ||||| ..... |||||
819 ATATGCTTAATGAGCTAAATTACTTATGAAATTAATAAACAAATGATGG 770
   ||||| ..... |||||
60 rOlYsIleuAspAspThrAspAsnLysGlyValLeu..... 71
   ||||| ..... |||||
769 GAAACATATATTGAGCTTGCTACTAAATGCTTAATGATATTTCATATT 720
   ||||| ..... |||||
72 .....AlaLeuGluLysAspIleAlaHis.....Se 80
   ||||| ..... |||||
719 AATGACACACTTGACATGCAATACAAATATAAATTCATCAATTAAATTA 670
   ||||| ..... |||||
80 rTyIleuAspThrValSerLeuIleProGlyAlaGluAsnLeuLeu.... 95
   ||||| ..... |||||
669 ATATGCTCAACAGCTAAACCAATACAGAGCAATGACGATTTTAATTT 620
   ||||| ..... |||||
96 .....LeuSerLeuAspValAspThrGluThrGlnLys 106
   ||||| ..... |||||
619 ATTTTAAAAAATCATCATCTTCTTACCTTCTGCTACTTCAACCATGAAA 570
   ||||| ..... |||||
107 LysLeuProGluArgLysTrp..... 113
   ||||| ..... |||||
569 TCTGCTCTTTGACAAAATAATGCTTAAACCAAGAGATGCTCAATTATTT 520
   ||||| ..... |||||
519 TGATGCAATTGCTCTTGAGATGATCTCAT..... 489
   ||||| ..... |||||
129 LuthrIleLeuLysAsnValGlyLysProLysValPheIleThrGlyPhe 145
   ||||| ..... |||||
489 ..... 489
146 AspValLysAsnGlyLysProAspProGluGlyTyrSerArgAlaArgAs 162
   ||||| ..... |||||
488 .....GTAAAGAGAGCTAAACCAATTCACAAATATTTTACATGCTGCTCA 442
   ||||| ..... |||||
162 pLeuLeuArgGlnAspLeuGlnLeuThrGlyLysGlnAspLeuLysTyr 178
   ||||| ..... |||||
441 TTTACTT.....GCATGCTACTGACATGAACACAG 413
   ||||| ..... |||||
179 ..ValValIlePheGluAspAlaProValGlyIleLysAlaGlyLysAlaMet 194
   ||||| ..... |||||
412 CAATGTATTTGAAAGATGACTGCTGCTTCAAGCAGCAATTCATCT 363
   ||||| ..... |||||
195 GAlaAlaIleThrValGlyIle 201
   ||||| ..... |||||
362 GAGAGCTTATCTGATGCTATT 342
   ||||| ..... |||||

seq_name: gb_est30:AU093031
seq_documentation_block:
LOCUS AU093031 635 bp mRNA EST 14-JUN-2000
DEFINITION AU093031 Rice callus Oryza sativa subsp. japonica cDNA clone C61715
, mRNA sequence.
ACCESSION AU093031
VERSION AU093031.1 GI:8528216
KEYWORDS EST
SOURCE Oryza sativa subsp. japonica.
ORGANISM Oryza sativa subsp. japonica
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

```

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Erihartoideae; Oryzeae; Oryza.
1 (bases 1 to 635)
Sasaki, T. and Yamamoto, K.
Rice cDNA from callus (2000)
Unpublished (2000)
Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai, Tsukuba
Ibaraki,
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasaki@ab.affrc.go.jp
PROJECT = "RGP".
C61715_82.

FEATURES
source .1..635
Location/Qualifiers
/organism="Oryza sativa subsp. japonica"
/strain="cultivar Nipponbare, sub_species japonica"
/db_xref="taxon:39947"
/clone="C61715"
/clone_lib="Rice callus"
/note="Vector: pBluescript II SK+; Site_1: SalI; Site_2:
NotI; cDNA prepared from rice callus mRNAs by using
oligo(dT) as a primer and ligating to the SalI-NotI site
of pBluescript II SK+ phagemid. 5 others
BASE COUNT 182 a 132 c 144 g 172 t

alignment_scores:
Quality: 117.50 Length: 82
Ratio: 2.350 Gaps: 1
Percent Similarity: 60.976 Percent Identity: 35.366

alignment_block:
US-09-558-284-2 x AU093031 ..

Align seg 1/1 to: AU093031 from: 1 to: 635

```

140 ValPheIleThrGlyPheAspValLysAsnGlyLysProAspProGlu 156
   ||||| ..... |||||
51 CTGCTCGTAATCGGAGGAGCTGTGACGAGCAACCATTCCTCGATGCC 100
   ||||| ..... |||||
156 yTyISerArgAlaArgAspLeuLeuArgGlnAspLeuGlnLeuThrGly 173
   ||||| ..... |||||
101 TTACTTCAGAGCT.....CTTGAGCTTACTGCTG 129
   ||||| ..... |||||
173 ysgInAspLeuLysTyrValValPheGluAspAlaProValGlyIleLys 189
   ||||| ..... |||||
130 CATCACTGATCACACATTCATTTTGGAGACTCCGGCTCAGGAGATTCGA 179
   ||||| ..... |||||
190 AlaGlyLysAlaMetGlyAlaIleThrValGlyIleThrSerSerTyrAs 206
   ||||| ..... |||||
180 GCCGCTGACTGCAAAATGTTCTGCTGCTGCTGCACTAGGAATCC 229
   ||||| ..... |||||
206 pLysSerValLeuPheAspAlaGlyAlaAspTyrValValLysAsp 221
   ||||| ..... |||||
230 AGAGAAAGTTCTGCAAGATGAGGAGCAACGCTGCTGTAATAAGAT 275
   ||||| ..... |||||

seq_name: gb_est50:AW696074
seq_documentation_block:
LOCUS AW696074 655 bp mRNA EST 15-JUN-2000
DEFINITION NE102C08ST1F1065 Developing stem Medicago truncatula cDNA clone
NE102C08ST 5', mRNA sequence.
ACCESSION AW696074
VERSION AW696074.1 GI:7570836
KEYWORDS EST
SOURCE barrel medic.
ORGANISM Medicago truncatula

```

```

TITLE
JOURNAL
COMMENT
REFERENCE
AUTHORS
    Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta
    Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots:
    Rosidae: eurosoids: Fabiales: Fabaceae: Papilionoideae: Medicago.
    1 (bases 1 to 655)
    He,X.-Z., Shadle,G., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell
    ,C.J., Flores,H.R., Inman,J.T., Weller,J.W., May,G.D. and Dixon
    ,R.A.
    Expressed Sequence Tags from the Samuel Roberts Noble Foundation
    Medicago truncatula stem library
    Unpublished (2000)
    Contact: Dixon RA
    Plant Biology Division
    The Samuel Roberts Noble Foundation
    2510 Sam Noble Parkway, Ardmore, OK 73402, USA
    Tel: 580 221 7302
    Fax: 580 221 7380
    Email: radixon@noble.org
    Insert length: 655 Std Error: 0.00
    Plate: 102 Row: C Column: 08
    Seq primer: TCACACGAGAAACGCTAAGAC.
    Location/Qualifiers
        1. 655
            /organism="Medicago truncatula"
            /db_xref="taxon:3880"
            /clone_lib="NF102C08ST"
            /clone_1lb="Developing stem"
            /tissue_type="stem"
            /dev_stage="Pooled developmental"
            /note="Vector: Lambda Zap; Contains a mixture of
            internodal stem segments"
BASE COUNT      188 a      127 c      149 g      191 t
ORIGIN

alignment_scores:
    Quality: 117.00      Length: 205
    Ratio: 1.093      Gaps: 9
Percent Similarity: 52.195      Percent Identity: 26.829

alignment_block:
US-09-558-284-2 x AW696074 ..

Align seg 1/1 to: AW696074 from: 1 to: 655

10  lllleuaphasplleuaspglythrlllevalserthrthvalaialaajuly 26
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
96  TTGTTTGATATAGAGGAACTCTTGATGTGACCAATCCACTATTATA 145
26  salatrphrlyslencystyrygluylryglylaasp..... 38
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
146  TGCCTTTCGTAATGCTTCAAGAGATTGGTTTCATGGAGGCAATCTTA 195
39  ..Prosergluleuphelylshisser...Hislqylaaaythrgringlu 53
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
196  TAACAGAGAAATTTTATTGCGACTGCTTCGCAAGACAAATCATGTGAC 245
54  Valleunatgargphepferolysleuaspspthraspsllysllyya 70
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
246  ATTGCTTTGGATCTCTTCCCC.....GGTGAATCGGAGCGAGGTTTAAA 289
70  lleualauleuglulysasptllealahiserrytleuaaspthvalserl 87
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
290  GTTGTGTAAGATTAAGAACCCATGTTTGGAGATTATAGCAGCTGAGCAAC 339
87  euileploglyalaglualsnleuleuleuenserleuaspyalaspthrclu 103
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
340  TGAAGCCG.....TTGAATGCTTGAT.....AAA 365
104  Thrghlnlyslsleuprogluarq.....LysTrrPALalilevalthrs 118
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
366  GTGAGGAAATGGATTGAAGATCTGGATTGAACGCAAGCTGCGCATTTACCAA 415
118  rgllyserProtyrleualaahseserTirpbhegluthrllleululsasny 135

```

```

416 TGCTCCAGACCAATGCA.....GAACGTATCTCCAAAGC 453
135 aAGlylyProllyValPhe.....lIethrGlyPheaspVallys 148
454 TTGGCTCTCAGATTCTTCATGCTGTATATATATGATGAATGGAA 503
149 AasnGlylyProaspProGluGlyTyrSerArgAlaArgaspLeuuar 165
504 CATGCCAAGCCTCATGCCAAGACCTACTTGAAAGCTTTCAGAGCTGCA 553
165 gGlnAspLeuGlnLeuThrGlyGlnAspLeuLysTyrValValPheG 182
554 AGCATCAAGATCATCACAC.....TTATATATTG 582
182 lIuAspAlaProValGlylIleLysAlaGlyLysAlaMetGlyAlaIleThr 198
583 AGCATTTCTGTTTCAGAAATCAAACTGCTGTGCTGCAGGAGATGCTGTT 652
199 ValGlylIleThrSer 203
653 ATAGCTATATCTACT 647

seq_name: gb_est30:AU100758

seq_documentation_block:
LOCUS AU100758 715 bp mRNA EST 22-AUG-2000
DEFINITION AU100758 Rice callus Oryza sativa subsp. japonica cDNA clone Cl2953
, mRNA sequence.
ACCESSION AU100758
VERSION AU100758.1 GI:9865008
KEYWORDS EST.
SOURCE Oryza sativa subsp. japonica.
ORGANISM Oryza sativa subsp. japonica
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrharioideae; Oryzaceae; Oryza.
REFERENCE
1 (bases 1 to 715)
Sasaki,T. and Yamamoto,K.
Rice cDNA from callus (2000)
Unpublished (2000)
JOURNAL
Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai, Tsukuba
Ibaraki,
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasaki@abr.affrc.go.jp
PROJECT "RGP".
Cl2953_82.

FEATURES
source location/Qualifiers
1..715
/organism="Oryza sativa subsp. japonica"
/strain="cultivar Nipponbare, sub-species japonica"
/db_xref="taxon:39947"
/clone="Cl2953"
/clone_lib="Rice callus"
/note="Vector: pBluescript II SK+; Site_1: SalI; Site_2:
NotI; cDNA prepared from rice callus mRNAs by using
oligo(dT) as a primer and ligating to the SalI/NotI site
of pBluescript II SK+ phagemid."
BASE COUNT 209 a 152 c 158 g 194 t 2 others
ORIGIN

alignment_scores:
Quality: 115 50 Length: 90
Ratio: 2.100 Gaps: 1
Percent Similarity: 61.111 Percent Identity: 32.222

alignment_block:

```


THIS PAGE BLANK (USPTO)